

This electronic thesis or dissertation has been downloaded from the King's Research Portal at <https://kclpure.kcl.ac.uk/portal/>



**Constructivism and the question of Objectivity
Fichte's ethics as Critique of Kant's**

Kazarinov-Hawk, Kit

Awarding institution:
King's College London

The copyright of this thesis rests with the author and no quotation from it or information derived from it may be published without proper acknowledgement.

END USER LICENCE AGREEMENT



Unless another licence is stated on the immediately following page this work is licensed

under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International

licence. <https://creativecommons.org/licenses/by-nc-nd/4.0/>

You are free to copy, distribute and transmit the work

Under the following conditions:

- Attribution: You must attribute the work in the manner specified by the author (but not in any way that suggests that they endorse you or your use of the work).
- Non Commercial: You may not use this work for commercial purposes.
- No Derivative Works - You may not alter, transform, or build upon this work.

Any of these conditions can be waived if you receive permission from the author. Your fair dealings and other rights are in no way affected by the above.

Take down policy

If you believe that this document breaches copyright please contact librarypure@kcl.ac.uk providing details, and we will remove access to the work immediately and investigate your claim.

This electronic theses or dissertation has been downloaded from the King's Research Portal at <https://kclpure.kcl.ac.uk/portal/>



Title: Constructivism and the question of Objectivity

Fichte's ethics as Critique of Kant's

Author: Kit Kazarinov-Hawk

The copyright of this thesis rests with the author and no quotation from it or information derived from it may be published without proper acknowledgement.

END USER LICENSE AGREEMENT



This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivs 3.0 Unported License. <http://creativecommons.org/licenses/by-nc-nd/3.0/>

You are free to:

- Share: to copy, distribute and transmit the work

Under the following conditions:

- Attribution: You must attribute the work in the manner specified by the author (but not in any way that suggests that they endorse you or your use of the work).
- Non Commercial: You may not use this work for commercial purposes.
- No Derivative Works - You may not alter, transform, or build upon this work.

Any of these conditions can be waived if you receive permission from the author. Your fair dealings and other rights are in no way affected by the above.

Take down policy

If you believe that this document breaches copyright please contact librarypure@kcl.ac.uk providing details, and we will remove access to the work immediately and investigate your claim.

Christopher Kazarinov-Hawk

MPhilStud

Constructivism and the question of Objectivity: Fichte's ethics as Critique of Kant's

Table of Contents.

<i>Abstract</i>	p.3
<i>Introduction</i>	p.3
§1: <i>Intelligible Entities and Things in Themselves</i>	p.4
§1:i) <i>The Third Antinomy: Exegesis</i>	p.5
§1:ii) <i>Intelligible Entities</i>	p.9
§1:iii) <i>Things in Themselves</i>	p.13
§1:iv) <i>The Later Works</i>	p.22
§1:iv.i) <i>Groundwork</i>	p.22
§1:iv.ii) <i>The Critique of Practical Reason</i>	p.23
§1:iv.iii) <i>Critique of Judgement</i>	p.26
§1:v) <i>Kantian Replies.</i>	p.32
§2: <i>Kantian Constructivism</i>	p.41
§2:i) <i>Constructivism and Metaphysics</i>	p.42
§2:ii.i) <i>Rawls</i>	p.43
§2:ii.ii) <i>O'Neill</i>	p.45
§2:ii.iii) <i>Korsgaard</i>	p.52
§2:iii.i) <i>Kant and Immediacy</i>	p.54
§2:iii.ii) <i>Kant and Conscience</i>	p.60
§3: <i>Fichte's Ethics</i>	p.67
§3:i) <i>A non-thought</i>	p.68
§3:ii.i) <i>Freedom without the thing in itself</i>	p.70
§3:ii.ii) <i>Practical Concerns</i>	p.74
§3:iii) <i>Intersubjectivity</i>	p.76
§3:iv) <i>Fichte and Conscience</i>	p.79
<i>Conclusion</i>	p.84

Abstract

Central to Kant's moral philosophy are the notions of autonomy and spontaneity, and their relation to reason and the understanding. Recent 'constructivist' readings of Kant's ethics thus emphasise the role of the subject's *reflection* in moral actions - reason is the only guarantor of the moral, and the right action must be worked out by the subject and consciously assented to. In contrast, for Fichte the moral is simply self-evident and immediately known to the subject. If Kant views the moral as requiring reflection and Fichte views the moral as immediate certainty, then it seems at first glance that the two are at loggerheads. Yet Fichte regarded himself as completing Kant's Critical project by simply following through Kant's thought to its fullest conclusions. Rather than dismissing Fichte's claim to complete Kant's philosophy, I suggest that paying close attention to Kant's ethics reveals him to be closer to Fichte than is often recognized.

Introduction

In the first chapter I question Kant's suggestion that a distinction between the thing in itself and appearances is helpful in moral matters. I suggest that Kant finds two modally distinct uses for the thing in itself, and that this is as problematic for Kant as it is for modern Kantians. Given that Kant's appeal to the thing in itself is ill-suited to serve its purposes in securing his practical philosophy, in §2 I discuss some recent 'constructivist' interpretations of Kant's ethics that attempt to do away with thing in itself whilst ostensibly salvaging the core insights of Kant's moral theory. I suggest that constructivist interpretations tend to offer a corrupt account of the phenomenology of morality by ignoring its immediacy in experience. This is particularly unfortunate since it is something that Kant was at pains to integrate into his moral writings. Indeed, in departing from Kant's careful insistence to preserve moral phenomenology I suggest the constructivists reveal their mis-reading of Kant's moral project. In §3 I conclude with a discussion of Fichte's ethics, focusing on his attempt to retain Kant's focus on moral phenomenology, but without introducing the thing in itself.

§1: Intelligible Entities and Things in Themselves

Consider Kant's famous claim, the 'positive' side of the critical project (Bxxv), that only by introducing a distinction between the thing in itself and appearances can the appropriate (that is, non-dogmatic) space be made for the possibility of the freedom necessary for morality (Bxxvii-xxx, A532/B560-A558/B586). Kant's claim and the argument to back it up both exploit the non-contradiction of the *idea* of freedom with strict causal determination, given transcendental idealism's commitment to the distinction between appearances and the thing in itself. My query in this opening chapter, expressed broadly, concerns not the issue of non-contradiction between the idea of freedom and causal determination, but rather, the assumption of coherence between Kant's claims regarding freedom and causality on the one hand, and on the other, transcendental idealism's commitment to 'the twofold meaning' (Bxxvii) brought about by the appearance/thing in itself distinction – why should we think that a distinction between appearances and the thing in itself would help us in inquiries into *moral* matters?

Since the bare bones of Kant's thoughts on this matter are presented first in the Third Antinomy and then fleshed out with more detail in the further two *Critiques* and *The Metaphysics of Morals* (hereafter *MM*), I turn first to the relevant issues as they are addressed in the Third Antinomy¹ and then look to the later works to see how Kant develops the idea.

¹ My exegesis of the Third Antimony draws heavily on the accounts offered by Allison (1990) and Bird (2006), and as such, ought to contain nothing controversial to readers familiar with the 'two aspects/points of view' interpretation of transcendental idealism. That is not to say that my argument assumes this reading to render Kant's thought coherent, or that it is the correct interpretation. Rather, I hope that this opening chapter puts pressure on Kant whether we read him as providing an alternative *to* traditional metaphysics/ontology (see, eg, Allison (1990), Bird (2006)), or as closer to traditional metaphysics/ontology (eg, Ameriks (2000), diGiovanni (2005)). I bracket discussion of the broader (perhaps more interesting) question of the actual metaphysical and ontological commitments of transcendental idealism.

§1:i) *The Third Antinomy: Exegesis*

The thesis and antithesis of the Third Antinomy respectively affirm and reject the possibility of our freedom, given the strict causal determination we find in our experience of nature, and to which Kant is apparently committed, by the Second Analogy.

Thesis: Causality in accordance with laws of nature is not the only one from which all the appearances of the world can be derived. It is also necessary to assume another causality through freedom in order to explain them.

(A444/B472)

Antithesis: There is no freedom, but everything in the world happens solely in accordance with laws of nature.

(A445/B473)

From the point of view of spatiotemporal appearances, causality is to be regarded as the schematized form of the necessary succession of events as in time (B233-4). If we consider our actions as exhausted by their description as spatiotemporal events and thus viewed within the chain of causality, then our actions must be determined by their preceding causes (A445/B473). This leaves no space for the assertion or logical coherence of a cause whose efficacy is *not* dependent of some prior one. Strict causal determination follows, which rules out the possibility of the relevant form of freedom which is conceived of *transcendentally* as the instigation of an absolute or *spontaneous* beginning to an ensuing causal chain. ‘Thus transcendental freedom is contrary to the causal law’ (A445/B474).

Yet, this conclusion is counter-balanced by another point of view which maintains the necessity of precisely this idea of an absolute or spontaneous source of an ensuing causal chain. If such an idea is rejected there would be an endless series of explanatory causes for any single event: ‘If...everything happens according to mere laws of nature, then...there is...never a first beginning’ (A446/B474). If there is never a first beginning

then there would be ‘no completeness of the series on the side of the causes descending one from another’ (A446/B474). This would be an affront to the completeness of explanation, and leads to contradiction, since ‘the law of nature consists just in this, that nothing happens without a cause’ (A446/B474). So non-contradiction and completeness of explanation - indeed, the very acceptability of the antithetical claims regarding the natural world as we experience it - demand the idea of an absolute or spontaneous beginning to an ensuing causal chain. This just is to say that explanation demands the idea of transcendental freedom (A448/B476).

On the face of it, the thesis and antithesis stand in irresolvable opposition to each other. As Allison succinctly sums up the problem: ‘the thesis and antithesis [can be] taken to claim, respectively, not everything takes place according to mechanistic causality and everything takes place according to natural causality’ (Allison (1990) p.14). Kant then goes on to ostensibly resolve the Antinomy by reference to transcendental idealism’s distinction between phenomena and the thing in itself.

Recall that the categories only find their full legitimate application in appearances² (A147/B186-7, A236-7/B295-6), and that ‘phenomena’ is the generic term covering all appearances thought through the categories (A248-9) and thus rendered spatiotemporal. So, to say that the categories only have legitimate application in relation to appearances is just to say that they apply only to spatiotemporal phenomena. Recall also that ‘causality’ just is the schematized form of the representation of the necessary succession of events as in time (B233-4) which is to say, the relevant sensory representations have been organized by the understanding via the pertinent category. Causality then, *qua* schematized category, only finds legitimate application in relation to spatiotemporal phenomena, or, in general terms, to appearances.

Recall also that we know from the Aesthetic that it makes no sense to say that anything is *intrinsically* or *in itself* spatiotemporal, since time and space just are the pure forms of

² Of course, the categories do have *some* residual, merely ‘logical significance’ (A147/B186) aside from sensation, but not of the kind that yields knowledge. See p.14-15 below.

intuition belonging to sensibility (inner and outer) (A42/B59). So whatever else we might say about the thing in itself, neither space nor time can be legitimately attributed to it. But note then, that if causality only finds legitimate application in relation to spatiotemporal phenomena, and the thing in itself, however construed, cannot, by definition, be legitimately attributed either spatial or temporal extension, that causality cannot have any legitimate application in relation to the thing in itself.

Returning now to the Third Antinomy, and recalling that the antithesis maintained that strict causal determinism must be true because our actions must be determined by their preceding causes (A445/B473), we now see that this is true of our actions *only if* our actions are exhausted by their description as spatiotemporal events. Since causality is only applicable to spatiotemporal appearances, there is, logically at least, another contender to rival appearance's claim to an exhaustive explanation of actions and events, namely, the thing in itself. But of course, this is to introduce an element of contingency regarding determinism which *qua* contingency is enough to make remove the logical contradiction between the idea of transcendental freedom and strict causal determinism. If causality only holds for appearances, even if it holds for all appearances, so long as there is no logical contradiction in postulating the idea of the thing in itself (whatever that might amount to) as well as phenomena, then there is no logical contradiction in postulating both the idea of freedom and determinism. The competing claims are, given transcendental idealism's distinction between phenomena and the thing in itself, not logically incompatible. As Allison puts the point: "it becomes possible that both sides may be correct: the thesis, with its assertion of ...[a] transcendently free cause outside of experience; the antithesis, with its refusal to admit such a cause within experience' (Allison (1990) p.23-4). So, a provisional model of the resolution sketched thus-far could be drawn up something like this:

Antinomy:

strict causal determination <≠> idea of transcendental freedom/spontaneity

Resolution:

i) strict causal determination <=> spatiotemporal phenomena/appearances

*ii) idea of transcendental freedom/spontaneity <=> idea of non-spatiotemporal
thing in itself*

...so long as i) and ii) are (in some sense) distinct.

In the Third Antinomy, Kant fleshes out the distinction between phenomena and the thing in itself in three ways. At A534/B562 he contrasts the animal power of choice: ‘*arbitrium brutum*’, with the human power of choice: ‘*arbitrium sensitivum*, yet...*liberum*’, stressing that though, like animals, humans are subject to sensuous impulses, unlike animals, we are not necessarily compelled or coerced by them – our actions may *follow* natural effects, but it is far from clear that our actions arise exclusively out of natural effects. Kant also exploits the contrast in our everyday language between actual occurrence – that which *is* – and moral judgement or prescription – that which *ought to be* (A547/B575-A549/B577, A554/B582-A555-B583). I take it that the idea here is to bring about an awareness that our ordinary conception of obligation already presupposes, indeed *requires*, an appeal to the relevant idea of transcendental freedom or spontaneity. Postponing detailed discussion of this crucial ‘is/ought’ distinction for later, the discussion which I want to focus on here, and which runs almost right through the formal resolution, relies on a distinction between the empirical (which we might just call the spatiotemporal) and the intelligible. Briefly, I suggest that since the intelligible does not seem to be strictly reducible to or identifiable with *either* phenomena or the thing in itself, it is not clear *why* the distinction between phenomena and the thing in itself gives us insight or help in *moral* matters.

§1:ii) *Intelligible Entities*

The terminology ‘intelligible entities’ is not Kant’s, but I use it to bracket together the various uses and applications of ‘the intelligible’ that Kant employs in the Antinomy: intelligible cause/causality (A537/B565, A544/B572); intelligible object (A538/B566); intelligible faculty (A545/B573); intelligible ground (A545/B573) and intelligible character (A541/B569). As I see it, although Kant runs them together, there are two distinct lines of discussion employed in the Antinomy regarding intelligible entities. The first is found at A537/B565-A541/B569, in which *the thing in itself* is operative in resolving the antinomy by making the relevant space for intelligible entities, whereas in the second discussion found at A551/B579-A554/B582, *reason* seems to do that work. I address the latter discussion first.

The latter discussion begins by inviting the reader to suppose that reason *were* to have transcendental freedom, *ie* the capacity to instigate an ensuing causal chain, and that the ensuing chain would be spatiotemporal. ‘[R]eason’s action then could be called free...[and] the empirical character [would be]...determined in the intelligible character’ (A551/B579). The potential problem facing this hypothetical scenario, namely, that we are not acquainted with the intelligible, but only with the empirical/spatiotemporal, is met with the observation that the intelligible is at least ‘indicated through appearances’ (A551/B579). Presumably Kant has in mind here something like the ‘ought/is’ distinction mentioned at A547/B575-A549/B577 to back up this claim – if the empirical is exhausted by the ‘is’, then, it is presumed, the ‘ought’ which we encounter in experience must have its locus in some other arena. To the further problem of how it *could be* that, given strict causal determination of the spatiotemporal, reason could conceivably be such that it could usher in an absolute beginning to an empirical causal chain, the now familiar answer is offered:

...if reason can have causality in regard to appearances...reason is not sensible...
[and] there takes place here what we did not find in any empirical series: that the **condition** of a successive series of occurrences could itself be empirically

unconditioned. For here the condition is **outside** the series of appearances (in the intelligible) and hence not subject to any sensible condition or to any determination of time through any passing cause... [Yet] the very same cause in another relation also belongs to the series of appearances.

(A552/B580)

That is, for reason to be transcendently free, it would have to be ‘a mere intelligible faculty’ that as such, would ‘not be subject to the form of time, and hence not subject to the conditions of temporal sequence’ (A551/B579). Thus, if reason *were* to be transcendently free yet causally efficacious in time and space, we would have to ascribe to the human being both an empirical and *intelligible* character, such that ‘[r]eason is thus the persisting condition of all voluntary actions under which the human being appears’ (A553/B581). Further, the empirical character would have to be considered the ‘sensible schema’ (A553/B581) of the intelligible, and ‘every [empirical] action, irrespective of the temporal relation in which it stands to other appearances, [as] the immediate effect of the intelligible character of pure reason’ (A553/B581). We may summarize that the two important intelligible entities at play here are ‘intelligible character’, which makes possible the idea of ‘intelligent causality’, and that *here* both are accommodated by the non-empirical status of reason (A551/B579). That is, given that the empirical is exhausted by the ‘is’, the ‘ought’ is accommodated by intelligible entities, the locus of which, at A553/B581 is identified as *reason*.

Note then, that if we were to draw up a similar model as before for *this* discussion, *reason’s intelligible entities* would play the operative role in enabling the resolution of the Antinomy, and *not* the thing in itself:

The 'reason' model of the Antinomy:

strict causal determination <≠> idea of transcendental freedom/spontaneity

Resolution:

i) strict causal determination <=> empirical character viewed from standpoint of spatiotemporal phenomena/appearances

*ii) idea of transcendental freedom/spontaneity <=> idea of intelligible entities viewed from standpoint of non-spatiotemporal **REASON**.*

...so long as i) and ii) are (in some sense) distinct.

The key stages of the other, earlier, discussion of intelligible entities at A537/B565-A541/B569 run much the same as they do in the later one just rehearsed, but with the one important difference that there, the intelligible is viewed from the standpoint of the thing in itself, rather than from reason. Again, we learn that an intelligible cause would have to be 'outside the series' of spatiotemporal, empirical conditions, whilst its effects could only be 'encountered in the series' (A537/B565). And again too, we can attribute transcendental freedom only if we consider both an empirical and intelligible character (A539/B567), though here:

The first [empirical character] one could call the character of such a thing in appearance, the second its character as a thing in itself.

(A539/B567, my emphasis)

And of course this yields the following model:

The 'thing in itself' model of the Antinomy:

strict causal determination \nleftrightarrow idea of transcendental freedom/spontaneity

Resolution:

i) strict causal determination \Leftrightarrow empirical character viewed from standpoint of spatiotemporal phenomena/appearances

*ii) idea of transcendental freedom/spontaneity \Leftrightarrow idea of intelligible entities viewed from standpoint of non-spatiotemporal **THING IN ITSELF**.*

...so long as i) and ii) are (in some sense) distinct.

It seems clear that Kant (and with him Allison and Bird too) does not feel this distinction troubling enough to be worthy of serious comment. Within the discussion of the resolution, reason and the thing in itself are, if not conflated, at least seen as unproblematically interchangeable at stage ii) of the last two of the above models of the Antinomy. Yet it is an important distinction to make. On the 'reason' model, the relevant distinction that enables the possibility of the freedom necessary for morality is the distinction between spatiotemporal appearances and the faculty of reason, whereas it is the distinction between the thing in itself and spatiotemporal appearances that does this work on the model immediately above. And this raises the following question: given that it seems sufficient for the possibility of the freedom necessary for morality to contrast phenomena with reason (from the 'reason model' offered at A551/B579-A554-B582), why do we *also* need to contrast phenomena with the thing in itself? Or, to put it another way, why should we think that a distinction between phenomena and the thing in itself would help us in inquiries into *moral* matters, since it seems the distinction between phenomena and reason will suffice?

§1:iii) *Things in Themselves*

I take it that the two most pressing answers to the question just posed would be something like:

i) perhaps it is too much to *equate* reason with the thing in itself or to use them interchangeably in the resolution to the Third Antinomy, but reason at least shares certain important theoretical attributes with the thing in itself, such as the conceptualizability of ‘being outside of’ the spatiotemporal series (in the sense discussed above). In this sense, the thing in itself serves the theoretical purpose of revealing to us otherwise opaque distinctions about reason. In short, the thing in itself and reason are suitably close in important theoretical respects, so that the comparison, if not the conflation, of the two yields a theoretical bounty, of which, the discussion found in the resolution of the Third Antinomy stands as a prime example.

ii) the first, ‘thing in itself’ discussion at A537/B565-A541/B569 should be viewed as an extremely abstract discussion, such that lays down the necessary logical conditions for *any thing* to be transcendently free. The later ‘reason’ discussion at A551/B579-A554-B582 ought then to be read as a more specific discussion, laying down the necessary conditions for an intellect such as ours to be transcendently free. The first, ‘thing in itself’ discussion would then establish what is necessary for transcendental freedom generally, and in the second, ‘reason’ discussion those findings are hypothetically applied to a particular example, namely, human beings.

With these two responses in mind I turn to a discussion of the thing in itself, and suggest that Kant simultaneously runs two modally distinct uses of the thing in itself and that this is problematic because these two uses pull in opposite, indeed, *irreconcilable* directions. The result is that once again we are left asking why the distinction between phenomena and the thing in itself is helpful in moral matters.

An excellent appraisal of this issue, highlighting the divergent uses to which Kant puts the phenomena/thing in itself distinction to work can be found in diGiovanni ((2005) *esp.*

p.6-16; p.154-186)³. In contrasting the role of the thing in itself in the first two Antinomies with the use to which it is put in the Third, diGiovanni clearly elucidates Kant's distinct and incompatible conceptions of the thing in itself (diGiovanni (2005) p.153-5). Consider the First Antinomy, the crucial tension of which can be summarized as whether or not the world has a beginning in time and spatial limits (A426/B454, A427/B455). As in the Third Antinomy, the resolution of the First turns on the ideality, and thus limited scope of the spatiotemporal. Since things in themselves by definition are not intuitable, they cannot be schematized and so neither space nor time is legitimately attributable to them. Yet in the First Antinomy this now familiar line is put to work *not* to reveal the merit in each of the respective theses (as it is in the Third), but rather, to bring about an awareness that the questions raised by the thesis and antithesis are simply *empty of meaning*. The validity of the two opposing theses of the First Antinomy is denied by the invocation of critical ignorance: if the spatiotemporal is ideal, then to be spatiotemporal just means not to have any *intrinsic* spatial or temporal extension (recall A42/B59 and p.8 above), but rather means that such limits depend on an external observer. So to ask about the *absolute* spatiotemporal character of anything is to misunderstand the parameters of meaningful inquiry for finite rational beings such as ourselves. Things in themselves are such that we could not possibly know them, and so in response to the sort of questions raised in the First Antinomy, 'one can answer that the **question** itself **is nothing**, because no object for the question is given' (A479/B507n).

Thus we cannot evade the obligation of giving at least a critical resolution of the questions of reason before us by lamenting the narrow limits of our reason and confessing, with the appearance of a modest self knowledge that [they] lie beyond our reason ... For each of these questions concerns an object that can be given nowhere but in our thoughts
(A481/B509)

³ I draw heavily on diGiovanni (2005) in this section. I am greatly indebted to his discussion. His insight into the contrary uses to which Kant employs the thing in itself forms the bedrock for my ensuing discussion of the inappropriateness of the thing in itself as locus for intelligible entities.

The thought that underwrites the resolution of the First Antinomy thus seems to be something like: the idea of the thing in itself stands in contradistinction to our experience, and so makes possible the comprehension of the limits of reason, and thereby also, the limits of our sphere of *meaning*. To briefly address Kant's theory of meaning, for Kant, when we get the full picture of schematised categories enabling successful, unified experience of objects, those concepts have full 'determinate' meaning [*Bedeutung*] (A245). This point's importance warrants explicit statement: fully determinate meaning for *a priori* concepts in experience is dependent upon their combination with intuition in the formation of judgements about objects (B298). Though Kant's most detailed discussions of 'meaning' appear in the Refutation, Schematism and Phenomena and Noumena chapters (see especially B186-7, A244-5, B298, and B303-5) perhaps the clearest way to bring out Kant's position is by reference to his famous comment: 'Without sensibility no object would be given to us, and without understanding none would be thought. Thoughts without content are empty, intuitions without concepts are blind' (A51/B75). Kant's position regarding 'meaning' is simply an extension of this thought. The categories have significant use only insofar as they are related to possible experience, otherwise they are simply 'empty of content' (B298). Full determinate meaning for concepts is found only when sensibility and understanding come together to make right the respective deficiencies of blindness and emptiness, enabling possible experience of objects in space (B186). The further thought is that though it still makes sense to talk of the pure, unschematised categories aside from their role in determinate fully meaningful judgements, such utterances would be 'empty', finding only a negligible 'logical' or 'functional' meaning (B186-7). Thus to speak of the thing in itself in categorial terminology is not meaningless in the sense that such talk of the thing in itself is not gobbledygook. That is, we understand the way such terms might be related 'functionally' or 'logically', and it is for that reason that such talk does not strike us as sheer gibberish. Nonetheless, there is another sense in which, for Kant, such talk *is* meaningless – lacking 'content' such talk is vacuous since it is ultimately *empty*⁴. Such talk lacks the sort of meaning that is valuable to us as human knowers, *ie*, full

⁴ See Kant's own example of the 'great wind' that surrounds 'metaphysically-great men' (*Prolegomena*, 373n).

determinate meaning. So whilst phenomena can be said to have full meaning for us, the thing in itself thus conceived can have at best only a logical significance, empty of meaning, and thereby stands to delineate the limits of possible knowledge and legitimate inference for a finite consciousness such as ours. Turning to diGiovanni:

[w]hether the cosmos is infinitely extended in space or has spatial limits, or again, whether it has existed for an infinite period of time or has rather a beginning in time, are questions that Kant claims cannot be answered, not because we have no sufficient knowledge for dealing with them, but because they are in principle unanswerable. They are meaningless.’

(diGiovanni (2005) p.154-5).

In contrast, in the Third Antinomy the idea of the thing in itself makes possible the idea of transcendental freedom, given strict causal determination of the spatiotemporal. And diGiovanni is correct to point out that these are distinct notions and applications of the thing in itself – the thing in itself is being put to work to do two very different jobs. In the First Antinomy what is made possible by the introduction of the idea of the thing in itself is an awareness of the limits of the sphere of meaning for finite consciousness, whereas in the Third, what is made possible is the idea of a spontaneous beginning of an ensuing causal chain. Yet this raises the question, if for beings like us, the meaning ‘buck’ stops within appearances, and, if causality falls squarely within that sphere or space of meaning (from the Schematism and the Second Analogy), what would the postulation of a causality aside from appearances amount to?

It is important to note that the first two Antinomies are formally distinguished from the latter two as mathematical, not dynamical, and so we should not expect the same form of resolution for the two different forms of antinomy⁵. Yet diGiovanni’s insight is still an

⁵ diGiovanni himself deals with this issue not under the narrow parameters of the forms of antinomy, but amongst the broader topic of the (mis)match within the Critical project’s elucidation of ‘intentional activities’ (diGiovanni (2005) p.160). Thus, ‘[g]ranted Kant’s purely formal conception of reason, there is still a conceptual obligation to define how the two sets of intentional activities in which reason engages (the practical and the theoretical) relate to one another and respect each other’s boundaries within the same mental universe’ (diGiovanni (2005) p.161-2). Whilst I will ultimately argue for a similar position

apt one - the core of the issue here is precisely that there *are* two different types of antinomy at play here, and that each finds a distinct form of resolution in the thing in itself. Since the thing in itself performs two different roles in resolving the two different forms of antinomy, the pressing question is whether these two different jobs can co-exist in the same construct. That is, can the thing in itself be coherently conceived as performing *both roles*?

I mean here not simply to repeat the old suggestion that it is only by an equivocation in the use of the term ‘causality’ (*ie* intelligible vs. empirical) that the Third Antinomy is resolved⁶. I mean to go further and suggest that if, as the resolution to the First Antinomy suggests, the limit of our sphere of meaning is marked by the thing in itself, then from the point of view of the thing in itself there must simply be an empty, blind *IS*, void of meaning – recall the criteria for emptiness and blindness at A51/B75. And of course from that point of view, it is not only *causality* that finds illegitimate application, but, crucially, the all-important *OUGHT* of the Third Antinomy’s intelligible entities must surely also be condemned to the same fate. That is, the strength of the ‘*ought*’, as distinct from the ‘*is*’, lies precisely in the fact that it is thoroughly *saturated* with meaning, not void of it. So, if the foregoing is correct, and at least at some points (*eg* the First Antinomy) Kant’s conception of the thing in itself delineates the boundaries of our sphere of meaning, *ie*, that there is only a blind *is* from the point of view of the thing in itself, then the required ‘*ought*’ of the Third Antinomy’s intelligible entities, cannot be coherently viewed from that blind ‘*is*’ point of view. Indeed, we might say that Kant is disingenuous in suggesting that the contrast between the thing in itself and phenomena ushers in a ‘twofold meaning’ (Bxxvii, and p.4 above) for the Critical philosophy. Rather, the lesson of the First

regarding the mismatch of theoretical and practical reason, diGiovanni’s particular framing of the issue seems inseparable from his wider stance regarding the metaphysical/ontological commitments of transcendental idealism, a topic I have postponed for a later paper. Thus, I have here chosen to limit my discussion to the forms of antinomy and the respective uses found for the thing in itself, and will follow a ‘metaphysically neutral’ path toward a discussion of the wider mismatch of Kant’s use of theoretical and practical reason.

⁶ See *eg* Paton (1946), and Allison (1986) p.398 (though Allison later retracted the accusation of an unjustified equivocation from spatiotemporal causality to freedom – see Allison (1990), p.244, and especially p.244n25). See also Wood “Kant’s Compatibilism” in Wood (1984) p.73-101, *esp* p.87-9 for a defence of Kant.

Antinomy is that the thing in itself and meaning are *modally distinct*. Returning again to diGiovanni:

Kant was saying that from that standpoint [of the thing in itself] all the distinctions that define rationality no longer hold... It is to imply...that the fiction of assuming the standpoint of the thing in itself is just a rhetorical device, since the concept of 'knowledge' no longer applies from that standpoint."

(diGiovanni (2005) p.20).

With diGiovanni's insight now in view, we see just how ill-suited the thing in itself is to accommodate intelligible entities. Recall from the preceding discussion that Kant claimed the intelligible entities cannot be coherently conceived of from the point of the view of phenomena. Now though, it seems that if intelligible entities can be coherently conceived of from the point of view of the thing in itself *at all*, they can do so only by virtue of an equivocation in the conception and application of the thing in itself. Since intelligible entities - *qua intelligible* - find no place whatsoever in the blind '*is*' outside of, or perhaps better, *aside* from our sphere of meaning, then if we are to retain the link between intelligible entities and the thing in itself, we must surely be forced to conceive of *two* modally distinct viewpoints of the thing in itself: the first being the blind '*is*'; the other a viewpoint from which deterministic spatiotemporal appearances can be contrasted, such that makes possible the idea of transcendental freedom necessary for morality, but that is distinct from the blind *is*. More, *this* point of view is, as necessary for the possibility of morality, *saturated* with, rather than empty of meaning. Returning to the discussion of the Third Antinomy's use of the 'ought/is' distinction (A547/B575-A549/B577; A554/B582-A555/B583, and above, p.8), we can now comment that even if we admit, with Kant, that the 'ought' we find in moral experience cannot be accommodated by the empirical, it is *not* clear that Kant has found an unproblematic locus for it in the thing in itself. Kant is employing the thing in itself in two very different capacities, and it is far from clear that the two could match up. We might draw up a model of the present issue as:

i) Thing in itself ==> blind '*is*' beyond, or aside from human sphere of meaning

- ii) Thing in itself ==> sphere of non-spatiotemporal meaning within which fall the intelligible entities necessary for the idea of transcendental freedom.

If the foregoing is correct, then it seems that either we must accept an incoherence within the critical project, given an unpalatable equivocation in Kant's use of the thing in itself – conceived of at some points as entirely void of meaning, and at others, as saturated with meaning – or we might want to make the following, three-fold distinction regarding the thing in itself, intelligible entities and appearances, rather than the two-fold one that Kant makes:

- i) Thing in itself ==> blind *is*, beyond, or aside from sphere of human meaning
- ii) Appearances ==> spatiotemporal phenomena
- iii) Intelligible Entities ==> not phenomena, since not spatiotemporal, yet not the thing itself either, since saturated with human meaning.

Now, if we make this three-fold distinction and divorce intelligible entities from the thing in itself, and we recall that on the 'reason' model of the Antinomy outlined above it was not necessary to include the thing in itself in order to display its formal resolution, we are again faced with the question, why would we think that a distinction between the thing in itself and phenomena would help us with moral matters? Note though that the point now is not *only* that reason can *also* play that role, but further, that since intelligible entities seem not to fit squarely with the use Kant finds for the thing in itself in other places, yet they *do* seem to fit squarely within the use he finds for reason, that the 'reason' model seems *a better fit* for intelligible entities than the 'thing in itself' model. Consequently, Kant's distinction between the thing in itself and phenomena looks ever-more redundant, indeed, even unnecessary for the possibility of morality. We then seem to have increasing

cause to ask: if the distinction between human reason and phenomena will not only do the required work in enabling the non-contradictory conceptualization of the idea of transcendental freedom, but seems to do it better than (that is, less problematically than) the idea of the thing in itself, then why do we need to include the thing in itself in our moral theory at all?

With the above three-fold distinction between the thing in itself, phenomena and intelligible entities in mind, we can address the two points presented earlier, each of which defended Kant's use of the thing in itself. Recall the first suggestion, that the thing in itself and reason are suitably close in important theoretical respects, so that the comparison, if not the conflation, of the two yields a theoretical bounty, of which, the discussion found in the resolution of the Third Antinomy stands as a prime example. To this it is surely sufficient to reply that whatever the bounty might be *within Kant's theoretical enterprise*, with regard to his *moral* project, from the foregoing it is clear that the distinction between reason and phenomena does the necessary work better than that between the thing in itself and phenomena. Thus, at least with regard to the possibility of transcendental freedom relevant to his moral project, we can answer that the bounty is still reaped, nay, yields a better crop without the thing in itself. And of course, this brings us back yet again to the question: why then would we think that the distinction between the thing in itself and phenomena is of use to us in *moral* matters?

Recall also the second suggestion, that the discussion of the resolution found at A537/B565-A541/B569, from which the 'thing in itself' model was sketched, establishes what is necessary for transcendental freedom generally, and in the second, 'reason' discussion at A551/B579-A554-B582 those findings are hypothetically applied to a particular example, namely, human beings. The idea is that the invocation of the thing in itself is justified since it establishes the abstract logical conditions for freedom. To this, we can similarly reply that the operative distinction here is more fittingly made by the counterpoising of reason and appearances, since the 'reason' model better accommodates intelligible entities for the reasons already discussed. Yet we can also make a further comment, itself an extension of that thought. If it is correct to understand the idea of the

thing in itself as the idea of the blind ‘*is*’ that delineates the boundaries of meaning, then if Kant were to be attributing transcendental freedom to a *thing* aside from our experience of it, would he not be transgressing the limits of his own system? Surely the lesson of transcendental idealism’s distinction between phenomena and the thing in itself is that aside from our experience of them, *things simply have no intrinsic meaning*. Now since freedom falls squarely within the sphere of meaning, and meaning is relative to possible human experience, then to the question of the *intrinsic, absolute* freedom of a thing aside from possible human experience, we are surely entitled to reply: ‘the **question** itself **is nothing**, because no object for the question is given’ (A479/B507n). Is any other response, the resolution of the Third Antinomy included, not to ascribe meaning beyond the sphere of the meaningful? Indeed, if the idea of freedom itself falls so squarely within the sphere of human meaning, then what sense is there to be made of the idea of the freedom of a *thing in itself*?

Before moving on to consider some possible objections to this line of thought, I want to make good on the promise made earlier and look at how these issues play out in Kant’s later, explicitly practical works⁷.

⁷ Of course it ought to be noted that the issue at stake in the Third Antinomy is that of *transcendental* freedom – namely the instigation of an absolute or spontaneous beginning to an ensuing chain – whilst the issue principally at stake in the moral works is more often *practical* freedom – that is, the absence of necessitation by sensuous impulse. Yet since my argument focuses on the role and application of the thing in itself in discussions of *both* transcendental *and* practical freedom, I take it to be unproblematic to move from a discussion of the Third Antinomy to the later works.

§1:iv) *The Later Works:*

§1:iv.i) *Groundwork*

Towards the end of *Groundwork of the Metaphysics of Morals* (hereafter, *Gr*), in part III, Kant rehearses the central themes of the Third Antinomy, contrasting appearances with the thing in itself, such that we are invited to conceive of ‘a distinction, although a crude one, between a *world of sense* and the *world of understanding*’ (*Gr* 4:451). Further, there is explicit emphasis laid on the idea that these two ‘worlds’, or perhaps better, ‘standpoints’ (*Gr* 4:452) must be *conceived of* as meeting in the practical actions of finitely conscious agents such as ourselves (*Gr* 4:456). As we will see, within *Gr* Kant clearly thinks that it is the distinction between the thing in itself and appearances that makes the space necessary for morality – but since this entails conceiving of the standpoint of the thing in itself as inundated with meaning, it is problematic, for the reasons just discussed.

The discussion at *Gr* 4:451 begins with consideration of the empirical representations of the self yielded from inner sense, which offer only a partial insight into the self, namely, the self as appearance. ‘Even as to himself, the human being cannot cognize what he is in himself through the cognizance he has by inner sensation’ (*Gr* 4:451). Aware of the limited nature of this insight, the subject ‘must necessarily assume something else lying at their basis, namely his ego as it may be constituted in itself’ (*Gr* 4:451, my emphasis). This then yields the familiar picture of the two standpoints, here explicitly conceived of as standpoints from which we can view the subject:

...as regards mere perception and receptivity to sensations he must count himself as belonging to the *world of sense*, but with regard to what there may be of pure activity in him (what reaches consciousness immediately and not through affection of the senses) he must count himself as belonging to the *intellectual world*
(*Gr* 4:451)

With this conception of the subject at play, Kant then forms his model of free agency:

...he has two standpoints from which he can regard himself and cognize laws for the use of his powers and consequently for all his actions; *first*, insofar as he belongs to the world of sense, under laws of nature (heteronomy); *second*, as belonging to the intelligible world, under laws which, being independent of nature are not empirical but grounded merely in reason.

(*Gr* 4:452)

Note then that in *Gr* it is the conception of the I *in itself* that opens the door to the intelligible, *ie*, here Kant employs the ‘thing in itself model’ (rather than the ‘reason model’) as a way of accounting for free agency. The distinction between appearances and the thing (I) in itself enables the conceptualization of the intelligible entities necessary for the possibility of morality, which is to say that it is from the standpoint of the I *in itself* that the ‘*ought*’ is brought to the table. As such, the standpoint of the thing in itself, as used here, must be conceived of as a standpoint that is swamped with meaning, which, as we have seen, raises concerns regarding the overall coherence of the critical project, since in other places the standpoint of the thing in itself is conceived of instead as a standpoint entirely incapable of accommodating of meaning.

§1:iv.ii) The Critique of Practical Reason

Within *The Critique of Practical Reason* (hereafter, *C.Pract.R*). there are passages that clearly endorse at least some variant of ‘the thing in itself model’, most notably in a footnote at p.5/109 where Kant practically recites the same discussion just elucidated in *Gr* regarding the necessity of conceiving of the self *in itself*, and at p.105/231 where Kant comments that ‘if we were capable of...an intellectual intuition...we should perceive that this entire chain of appearances in regard to all that concerns the moral laws depends on

the spontaneity of the subject as a thing in itself⁸. Yet there are also a number of relevant discussions that bear a far greater resemblance to the ‘reason model’, where the pivotal distinction that makes the possible space for morality is that between sensible appearances and a non-sensible, intelligible ‘world’ given through *reason*, without the invocation of the thing in itself. Indeed, the common characterization of the principle aim of the second *Critique* is to demonstrate that reason has the capacity to sufficiently determine action *ie*, to show that pure practical reason *alone* can be moral. Thus the thing in itself finds no place in the initial formulation of the second *Critique*’s central question as posed in the introduction: ‘whether pure practical reason of itself suffices to determine the will’ (p.15/120), nor does it figure in many of the statements of the answer this question finds. That is, we are frequently reminded that it is *reason* that determines moral actions, once we allow the invocation of the now-familiar intelligible or ‘supersensible’ entities. And in many of these discussions there is no reference to the thing itself (*eg* p.24/134, p.31/142, p.34/145). Thus the three-fold model for understanding Kant’s moral picture suggested above – that of the distinct status of the thing in itself, appearances and intelligible entities – seems accommodated, even if only perhaps *unintentionally*, by *C.Pract.R.*

...the moral law...gives us a fact absolutely inexplicable from any data of the sensible world, and the whole compass of our theoretical use of reason, a fact which points to a pure world of the understanding...

This law (as far as rational beings are concerned) gives to the world of sense, which is a sensible system of nature, the form of a world of the understanding, that is, of a *supersensible system of nature*... [which] we might call...the *archetypal* world (*natura archetypa*), which we only know in reason (*C.Pract.R.* p.44/156-7, my emphasis)

⁸ See also p.5/109 and p.122-3/252, where Kant invokes the idea of the self as ‘noumenon in a world of the understanding’, conflating the intelligible with the noumenal, if not the thing in itself (*see also* p.50/164-5). I postpone discussion of what significance a distinction between noumena and the thing in itself actually amounts to for later, when I address Kantian replies.

I know myself *on the one side* to be an intelligible being determined by the moral law (by means of freedom), and *on the other side* as acting in the world of sense. It is the concept of freedom alone that enables us to find the unconditioned and intelligible for the conditioned and sensible without going out of ourselves. For it is our own reason that by means of the supreme and unconditional practical law knows that itself and the being that is conscious of this law (our own person) belong to a pure world of understanding, and moreover defines the manner in which, as such it can be active.

(*C.Pract.R.* p.112/238-9, my emphasis)

Further, *C.Pract.R.* also clearly contains the thought, later made more explicit in the third *Critique*, that what was above referred to as ‘the sphere of meaning’ is relative only to a finitely rational consciousness such as ours, and thus, that questions of absolute or *intrinsic* meaning are illegitimate. First, we learn that ‘notions... of a *motive*, of an *interest*, and of a *maxim*, can be applied only to finite beings’ (p.84/205). Then later, Kant invites us to imagine (though it is impossible) that we could experience things not just from our limited perspective, but rather, in their totality, such that ‘*God* and *eternity* with their *awful majesty* would stand unceasingly *before our eyes*’ (p.157/294). From that perspective, Kant adds, the many conflicts between our finite consciousness and the demands of reason would be avoided such that we would achieve a kind of mechanistic infallibility. Yet he also adds that in such a situation:

...the moral worth of actions, on which alone in the eyes of supreme wisdom the worth of the person and even that of the world depends would cease to exist. As long as the nature of man remains what it is, his conduct would be thus be changed into mere mechanism, in which, as in a puppet-show, everything would *gesticulate* well, but there would be *no life* in the figures.

(*C.Pract.R.* p.157/294, my emphasis)

The implication is clear – it is only by virtue of the kind of finite consciousness that we possess, and only within the sphere of possible experience *for* such a finite consciousness

that the sort of *meaning* and, now we can also add, *worth* necessary for morality can arise. Thus, to ask of the *intrinsic* morality of things aside from our experience is not to ask a legitimate question – intrinsically, things lack meaning and worth. In short, within *C.Pract.R.* we can find the thought, mentioned above in relation to the First Antinomy, that since it is, by definition, extrinsic to human experience, we can only expect a blind ‘*is*’ from the standpoint of the thing itself.

I take this to be sufficient to show that within *C.Pract.R.* Kant runs both the ‘reason’ and ‘thing in itself’ models concurrently in his account of the possibility of morality. Sometimes, as at p.5/109, the operative distinction is clearly between the thing itself and appearances, whereas at other times, such as at p.112/238-9, the operative distinction required to make space for the necessary intelligible entities is clearly that between reason and sensibility. Further, since Kant also runs the thought of the intrinsically blind ‘*is*’ of things aside of our finite consciousness of them, the ‘thing in itself’ model remains problematic for the reasons discussed above, namely that morality’s ‘*ought*’ surely requires the notions of *meaning* and *worth*, which as is here made explicit, are peculiar to the human standpoint. And of course this brings us back yet again to the question we have pressed throughout this opening chapter, namely, if Kant’s texts can be shown to accommodate (if not appreciate) the sufficiency of the ‘reason’ model to keep open the possibility of morality, and (as Kant seems not to appreciate at all), the ‘reason’ model in fact seems superior to the ‘thing in itself’ model in this regard, then why should we think that the thing in itself is of any use to us in *moral matters*?

§1:iv.iii) *Critique of Judgement*

A central theme of *Critique of Judgement* (hereafter, *C.ofJ.*) is that the only way we can make sense of the mutual inter-dependence of part and whole that we find in experience is through *our* introduction, indeed *imposition* of the distinctly *human* notion of purpose or ‘determination of ends’ upon intuition, such that the idea of the whole precedes and determines the production and arrangement of the parts. The comparison of a house with a product of nature at §65 (p.200-1/372-3) is a prime example.

Thus a house is certainly the cause of the money that is received in rent, but yet, conversely, the representation of this possible income was the cause of the building of the house...But if a thing is a product of nature and in this character is notwithstanding to contain intrinsically and in its inner possibility a relation to ends, in other words, is to be possible only...independently of the causality of rational agents, then this *second* requisite is involved, namely that... the idea of the whole may, conversely, or reciprocally, determine in its turn the form and combination of all the parts, not as cause – for that would make it an [human] art product – but as the ground for the cognition of the systematic unity of the form and cognition of the systemic unity of the form and combination of all the manifold contained in the given matter for the person judging it.

(*C.ofJ.* p.200-1/372, my emphasis)

As I understand it, the idea here is that natural purposes, such as we ordinarily experience (or perhaps better, *presuppose*) them, are not intelligible merely in terms of the laws of nature, and so the unity of experience presupposes that we impose upon intuition the notion of purpose, or purposiveness, analogous to that employed in the creation of, say, a house or an artwork. We see here the now familiar dichotomy in Kant's thought of, on the one hand, a conception of a blank-slate '*is*' of things aside from our experience of them, and on the other, *peculiarly human ideas*, of which 'purpose', or 'purposiveness' are examples, which we impose upon intuition, such that experience is unified and meaningful in the manner in which we find it. Indeed, this dichotomy finds its most explicit formulation within *C.ofJ.*, such that nature is regarded as 'mere mechanism', and it is 'only outside the conception of nature, and not in it, that we may hope to find some shadow of a ground...for that unity' (p.188/360).

Whilst the counterpoising of blind deterministic mechanism and meaningful purpose is largely familiar from the foregoing exegesis and discussion, there is a further thought in *C.ofJ.* that so securely cements the notion of the blind '*is*' of things aside from human

experience of them that it stands to distance Kant's two uses of the thing in itself even further, rendering them irreconcilable.

...if our understanding were intuitive it would have no objects but such as are actual. Concepts, which are merely directed to the possibility of an object, and sensuous intuitions, which give us something and yet do not thereby let us cognize it as an object, would both cease to exist. Now the whole distinction which we draw between the possible and the merely actual rests upon the fact that possibility signifies the position of the representation of a thing relatively to our concept, and, in general, to our capacity of thinking, whereas the actuality signifies the positing of the thing in its intrinsic existence apart from this concept. Accordingly the distinction of possible from actual things is one that is merely valid subjectively for human understanding. It arises from the fact that even if something does not exist, we may yet always give it a place in our thoughts, or if there is something of which we have no concept we may nevertheless imagine it given. To say, therefore, that things may be possible without being actual, that from mere possibility, therefore, no actuality may be drawn, is to state propositions that hold true for human reason, without such validity providing that this distinction lies in the things themselves.

(*C.ofJ.* p.229-30/402 my emphasis).

As I understand it, the thought here⁹ is that possibility signifies nothing about the content of a judgement *per se*, but denotes one of those curious contingencies that give rise to the sort of experience in which we find ourselves immersed. Possibility, as elucidated in the above quotation, is distinguished from actuality only as a consequence of the distinct cognitive capacities of the understanding and sensibility and their inter-relation in

⁹ It seems here pertinent to quote Schelling's remark that '[n]ever, perhaps, have so many deep thoughts been pressed together in so few pages as is the case in section 76 of the Critique of the Teleological Power of Judgement' (AA I,2:175; SWI:242, in Förster, E. (2012), p.141). There is a lot going on this short passage and, once again drawing heavily on diGiovanni (2005) (*esp.* p.21-24), I limit my discussion to the distinction made between possibility and actuality.

judgement – if intuition were *not* divorced from the understanding, that is, if the understanding were itself intuitive, possibility would fall since any distinction between abstract discursivity and the particular realization of objects in experience would simply disappear: we would see God and nature in all their awful majesty, but the possibility of conceiving otherwise would be lost. To distinguish between the possible and the actual presupposes the ability to distinguish what is given discursively from what is given from sensibility, which itself presupposes a cognitive faculty of understanding distinct from a faculty of intuition. Thus, possibility is unique to the human standpoint.

Yet if this is the case, and, as Kant puts it, ‘the distinction of possible from actual things is one that is merely valid subjectively for human understanding’, surely, from the *objective* viewpoint of the thing in itself, possibility finds no valid application at all. What we learn from §76 is that possibility arises only by virtue of the particular interaction of our limited cognitive faculties. Within this context, the thing in itself stands simply as a theoretical construct – a conceivable object that simply *is* and in regard to which possibility is entirely *alien* - the conceptual counterpart to anything we could ever experience, given the divorced, yet inter-related nature of the understanding and intuition in human consciousness. ‘[A]ctuality signifies the positing of the thing in its intrinsic existence apart from this concept.’ For Kant then, note the irony underwriting the human ability of conceiving otherwise: that by virtue of which we can conceive of the possible is also that by virtue of which we can conceive the possibility of the non-possible. Only by virtue of the interplay of the discursive and the receptive is it even conceivable to think of that which is given in sensibility aside from its interaction with the understanding – only by virtue of the concept of ‘possibility’ yielded by the understanding can one even conceive of what it would be like to lack an understanding, and thereby think the possibility of non-possibility.

In many ways, the thought expressed here simply fleshes out the implications of the ‘Copernican’ turn, as expressed at Bxvi, to investigate *not* whether cognition conforms to objects, but rather *whether objects conform to cognition*. The upshot of the critical investigation, as expressed at the end of the above passage (*C.ofJ.* p.230/402), is that the

nature of our cognitive faculties commits us to a particular kind of experience of the world, but that this peculiarly human experience to which we are committed *gives us no assurance* that there is anything in the objects *themselves* that corresponds to our experience. We have already seen that *meaning* (*CPR*), *worth* (*C.Pract.R.*) and *purpose* (*C.ofJ.*) all fall squarely within, even *constitute*, the uniquely human standpoint, beyond which they find no legitimate application. Now we learn that *possibility* too can be added to that list.

Now, I take it as a given that possibility is essential to any conception of morality. The entirety of our moral language, ‘responsibility’, ‘obligation’, ‘guilt’, ‘merit’, indeed, even the ‘*ought*’, all collapse unless we are committed to the *possibility* of things and events being other than they in fact are (or were). If possibility falls, it takes with it the whole artifice of judging actions against a norm that transcends their actuality. If that goes, it becomes inconceivable to coherently discuss (even if we fail to determine) how things *ought* to be. Without possibility, the *ought* not only fails to trump the *is*, it becomes a non-thought. As diGiovanni puts the point, ‘[i]f the distinction between actuality and possibility is strictly a function of human cognition, yet this distinction is essential to moral language, it follows that, when tested against the ideal thing in itself, moral language loses all meaning’ (diGiovanni (2005), p.21)¹⁰.

diGiovanni comments that Kant must be employing ‘possibility’ in two different ways, albeit unwittingly ((2005) p.21). On the one hand, Kant’s *theoretical* inquiries into the workings of our finite, limited consciousness leads him, as in §76, to view ‘possibility’ as a cognitive peculiarity of human beings. That is, at times Kant is committed to the strictly *relative* nature of ‘possibility’ – ‘possibility’ as unique to the human condition¹¹. Despite

¹⁰ Again, whilst diGiovanni takes a particular stance on what ‘ideal’ implies here, I leave that question unaddressed.

¹¹ It is on this basis that diGiovanni reads transcendental idealism as decidedly ‘metaphysical’ in the tradition of Descartes, Spinoza and Leibniz. Thus, “[t]his was however, as far-reaching a metaphysical commitment as any Enlightenment philosopher would have dared to advance’ ((2005) p.20) and also, “[s]o construed, [intellectual intuition in §76] is reminiscent of the theological notion of God’s creative act; or, closer to home, of Descartes’s and Spinoza’s *causa sui*” ((2005) p.22). Again, wishing to leave the question of the ultimate metaphysical status of transcendental idealism to one side, I have attempted to reconstruct a metaphysically neutral variation on diGiovanni’s theme, using his insight only as a point from which to critique the thing in itself (whatever its metaphysical baggage may be) as locus for intelligible entities.

this, indeed against this commitment, when it comes to Kant's moral theory, 'possibility' must surely hold beyond the exclusivity of the human standpoint, that is, *unconditionally*. The point is important not least because Kant famously tells us of the *unconditional* hold of the moral ought, but also because he introduces the kingdom of ends not as a community of finitely rational human being, but rather as a community of *purely rational beings* (Gr, 4:433).

With diGiovanni's further insight into the implications of Kant's uses of 'possibility' in mind, we see just how ill-suited the thing in itself is to accommodate the intelligible entities necessary for morality. Conceived of in the manner of §76, the thing in itself stands to *entirely negate* any real possibility of morality, rather than accommodating it. Indeed, if the thing in itself is here conceived of as the theoretical antithesis to possibility, each of the above two senses of possibility are negated from the viewpoint of the thing in itself. That is, if, for morality, possibility is to hold unconditionally (as in the kingdom of ends), and possibility and the thing in itself cannot be coherently conceived as coexisting in the same theoretical space, then the unconditional hold of possibility and the thing in itself simply fail to meet. Alternatively, if possibility holds only for human beings (and presumably then *not* in the kingdom of ends) it still finds no place from the viewpoint of the thing in itself. Thus, neither of the two conceptions of possibility, which Kant runs together, are accommodated by the postulation of the thing in itself. And note further, that yet again, the operative distinction here is between spatiotemporal appearances on the one hand - here viewed as that which is initially given through sensibility - and on the other, reason's faculty of the understanding. Granted, the intelligible entities necessary for morality and with them the uniquely human experiences of *meaning*, *worth*, *purpose* and *possibility* are nowhere to be found from the viewpoint of the *spatiotemporal*. Yet, they are nowhere to be found from this viewpoint of the thing in itself either. Thus, with the pivotal distinction necessary for the possibility of morality being reason and sensibility, and the thing in itself as a blind '*is*', an empty void equally as incapable as spatiotemporal appearances of accommodating the notions of meaning, worth, purpose and possibility necessary for morality, we are once again left asking: why should we

think that the distinction between the thing in itself and appearances will help us in moral matters?

§1:v) *Kantian replies.*

I have suggested that Kant runs two different conceptions of the thing in itself together within the critical project, and further, that because they are modally distinct they fail to meet up. Kant seems committed to envisioning the thing in itself as a blind *is* that stands as a counterpoint to our meaningful, moral lives and yet also employs the thing in itself as that which ostensibly makes such a life conceivable without contradiction. And of course, the problem is that the blind *is* just does stand to contradict the meaningful moral lives we live. Now, suppose we admit that the ‘blind *is* conception’ of the thing in itself is the result of Kant’s *theoretical* inquiries, indeed, marks the pinnacle of the investigation into *theoretical* reason. Suppose further, that the space for intelligible entities made by the thing in itself is the result of Kant’s *moral* inquiries, indeed, that this marks the pinnacle of the investigation into *practical* reason. Might a distinction between theoretical reason (and its impasse) and practical reason dissolve the concern about the divergent uses of the thing in itself? The materials for such a response might be thought to reside in Allison’s work.

Allison admits that theoretical (‘speculative’) reason fails to accommodate ‘the reality of an unconditioned causality’ necessary for the intelligible entities. He goes on to suggest that at best, theoretical reason can ‘merely defend it against the charge of inconceivability...by creating a “vacant place” for it in an intelligible world’ (Allison (1990) p.243). Allison’s further thought is that ‘pure practical reason fills this vacancy through the moral law, thereby resolving speculative reason’s problem... [Thus] the moral law gives a positive sense to the indeterminate notion of an intelligible causality.’ ((1990) p.244). The idea is that theoretical reason finds the end of its tether with the blind *is*, which, as we have seen, fails to accommodate intelligible entities, thereby bringing the possibility of morality into question. Nonetheless, we find ‘the ought’ in our moral experience. Theoretical reason thus simultaneously *requires* and *fails to accommodate*

that which is necessary for morality, and its inquiries into the thing in itself reflect as much. This is ok though, since practical reason makes good on theoretical reason's deficiency, by stepping into the 'vacant space' in demonstrating the initiation of intelligible causality in moral action (though the ultimate grounds of this intelligible causality remain unknowable, given theoretical reason's impasse). For Allison then, far from pulling in opposite directions, theoretical and practical reason are unified just in making right each others deficiencies. Practical reason demonstrates precisely that which theoretical reason needs but cannot find, and theoretical reason explains why its source is in principle unknowable.

The problem with Allison's line of thought is that it still fails to reconcile the two uses of the thing in itself at play. The heart of the problem is that theoretical reason does not find its end in a 'vacant space' *in the sense of* an empty room awaiting the delivery of some metaphysical or intentional furniture. Rather, theoretical reason is left concluding that the thing in itself must be a 'vacant space' in the sense of a *vacuum*, and a vacuum simply cannot accommodate *any furniture* – it is, by definition, simply a void. Allison appears to miss the importance of this modal distinction in commenting that *theoretical reason* defends intelligible entities 'against the charge of inconceivability'. In fact, theoretical reason leads us to the opposite conclusion from Kant's – Kant ought to conclude that unconditioned causality is not just inconceivable beyond the human standpoint, but is modally distinct from, and thus *not* attributable to the blind, empty thing in itself. It is not the case that theoretical reason *makes a space for* intelligible causality – rather, Kant ought to view intelligent causality as altogether distinct from theoretical reason's scope. As Strawson puts it: "[i]t has nothing to do with the interests of theoretical reason" (Strawson (1966) p.213).

More, it is for precisely this reason that Allison's insistence on reminding us that the idea of freedom plays a merely *regulative* role in Kant's philosophy (Allison (1990) p.45-6) also fails to get off the ground. Regulative or not, if the space made vacant by the theoretical philosophy is such that the demonstration of freedom in the practical philosophy cannot fit in it, then the two entirely fail to speak to each other. Worse, even if

Allison were correct to understand practical and theoretical reason as unified in moral acts, the puzzle over the thing in itself would still remain. Suppose we grant Kant (and Allison) that theoretical reason meets its exhaustion in the blind *is*, and further, that practical reason demonstrates precisely that causality that lies beyond this limit. Are we not still left asking why this intelligible causality finds accommodation in *the thing in itself*? Indeed, on Allison's account it is clearly (practical) *reason* that does the work, so why are we to go the extra step with Kant in postulating the thing in itself here? And if we do go that extra mile, surely we are still left head-scratching over how the thing in itself *qua* vacuum could possibly retain its (properly conceived) 'vacant' status whilst also accommodating intelligent entities – how could it be necessarily empty, yet pregnant?

Further, I suggest that once this point is properly appreciated, the force of the contemporary Kantian defensive arsenal is largely, if not entirely disarmed. Critical ignorance fails to dissolve this issue, as does timeless agency.

Kantians often like to maintain that criticisms of Kant's use of the thing in itself mistake its necessarily unknowable character for irresolvable difficulties or contradictions about Kant's position itself. Since, by definition, no knowledge can be attained of the thing in itself, it should come as no surprise that talk of it strikes us as not-a-little unfathomable and mysterious. What the Kantian philosophy offers is not a full explanation that removes this mystery or unfathomable-ness, but rather, the proper philosophical framework within which to articulate our ignorance. Thus, '[t]ranscendental idealism translates the mysteriousness of human freedom into something at least negatively comprehensible, giving our ignorance of its nature rational form' (Gardner (1999) p.264). So of course questions remain unanswered by the Critical philosophy, but that is because such questions are in principle unanswerable. Perhaps then the question of how the thing in itself can accommodate intelligible entities is a question of this kind. Freedom, Kantians might say, ought to remain weird for a consciousness like ours.

The obvious concern with the appeal to Critical ignorance is that insisting on the weirdness of the account might just mask some bad philosophy. As Freyenhagen puts it: “[t]he appeal to the black box of the intelligible realm does not give us sufficient assurance that what looks like a contradiction is no such thing.” ((2008) p.68). Indeed, the point at issue here is not that we want to know something beyond ‘the bounds of sense’. If it were, critical ignorance might be a good response. Instead the point is that the appeal to the bounds of sense so central to the theoretical project seems to preclude the possibility of the thing in itself being the locus of the intelligible entities necessary for the moral project. The appeal to Critical ignorance just does not bridge the distinct modalities of blind-ness and empty-ness on the one hand, and on the other the accommodation of the whole package of fully determinate meaning that Kant needs for the intelligible entities. What we have here just is a philosophically unpalatable equivocation from thing in itself *qua* vacuum when we need it in the theoretical project, to thing in itself *qua* empty room capable of accommodating philosophical decoration in the practical philosophy – and these two conceptions simply fail to hang together. So even if we grant the ultimate unknowability of the thing in itself, we are still left with a puzzle over what exactly it is that we are ignorant of – is it the blind void, or its modally distinct counterpoint, the locus of intelligible entities? Far from dissolving this issue then, critical ignorance is central to the problem.

More, any appeal to timeless agency here seems misguided. As we have seen, for Kant, the spatiotemporal arena cannot accommodate intelligible entities. Thus moral agency must be for Kant, in some sense *timeless* (A539/B567; A551-3/B579-81). Note though that whether or not Wood is correct to view the notion of a non-temporal agency as coherent given Kant’s understanding of causality (Wood (1983) p.86-9), or whether Allison is correct to think of agency as timeless in the sense that conclusions following from premises is timeless or even in the sense that the maxims underlying moral action are based on non-temporal principles (Allison (1990) p.47-53) – all of this is secondary to the question of *where* this non-temporal agency is supposed to be happening. That is, Kant locates this agency as operating at the level or viewpoint of the thing in itself, and the problem is that agency, *timeless or not*, requires a sphere of meaning, indeed some

accommodation of *meaningful* deliberation. Yet if agency requires meaning, and meaning is modally distinct from the thing in itself *qua* vacuum, then this understanding of the thing in itself is ill suited for Kant's moral project. We might say that for Kant, if the agency is timeless, then it must also be *meaningless*, since what there simply is no bridging the modal distinction he makes between the blind is on the one hand, and space, time and meaning on the other. Thus, the notion of timeless agency also fails to resolve the issue.

Indeed, as Freyenhagen (2008) has shown, with the case of timeless agency, we see a further, analogous complication with the thing in itself. From the aesthetic, we know that space and time are nothing but the pure forms of intuition, are thus ideal and that by 'ideal' it is understood that they do not pertain to the thing in itself (A32-6/B49-53). Now, if agency is to be remotely recognizable *as* agency it must surely involve the capacity to act, and the very notion of action is surely inseparable from the notion of time. There must be a 'moment' when principles are assented to, as even Allison admits. Yet, with 'moment', we are back speaking the language of time once more. As Freyenhagen puts it, 'on Allison's own reading...what grounds reason a free decision is an act of reason (the incorporation of an incentive into a maxim) and...this act has to be temporal' ((2008) p.83). So if timeless agency is accommodated only by the thing in itself, *and* is coherent only if time is postulated, then a new, separate conception of time is needed, if not a further temporal arena, aside from the deterministic spatiotemporal one that Kant has already dismissed. Further, since Kant has already made clear that time finds no application with regard to the thing in itself, if we are to accept that timeless agency is accommodated by the thing in itself yet still requires (new) time, it seems the account must end up looking worryingly un-Kantian.

Last, it will likely not have gone unnoticed that the discussion thus far has made no use of a possible distinction between the thing in itself and *noumena*. Indeed, the abundance of passages, spanning the entire critical enterprise, in which the attribution of intelligible entities to man is justified on the basis of the conceivability of man *as noumenon* (see *CPR* (A541/B569); *C.Pract.R.* (p.123/252); *C.ofJ.* (p.264/435) and *MM* (p.215/418))

would seem to add considerable credence to this thought. Further, Allison's comprehensive discussion of Kant's moral thought (1990) scarcely makes mention of the thing in itself *at all*, instead almost exclusively contrasting phenomena with *noumena*. Could such a distinction serve to alleviate the tension between the opposing uses of the thing in itself, such that the thing in itself is conceived of as the blind *is*, leaving space for the *noumena*/phenomena distinction to accommodate the intelligible entities necessary for morality? I suggest not, since noumena and the thing in itself are suitably close to each other in important respects that the criticisms leveled at the thing in itself hold true for noumena, even if the two can be distinguished from each other - but this will take some unpacking.

Firstly, Kant himself often simply assumes that there is no distinction to be made between the thing in itself and noumena. At one point he comments that we are free to equate to two (A288/B344-5), at other points he uses them interchangeably (A249), indeed, sometimes within the same sentence (A253, A254/B310, A259/B315). The two are undeniably part of a package since there is a considerable overlap between them. Yet subtle distinctions make it possible to distinguish them. The thought here is that the thing in itself can be taken to refer specifically to the idea of an object aside from it as given in experience. Recall the discussion of §76 of *C.ofJ.*, where actuality was said to signify 'the positing of the thing in its intrinsic existence' and thus void of the discursive elements that reason's faculty of understanding brings to it. In this sense, the thing in itself, *qua* idea of an object aside from how we represent it, stands only as a direct logical correlate to the cognition of objects in experience. Taken in this sense, the idea thing in itself arises out of philosophical reflection on perception and the status of the objects given therein. Noumena could then be distinguished from this narrow conception of the thing in itself if we were to understand by it the more general conception of a mere postulate of the pure understanding or reason (A249), the very idea of which arises *not* directly from meditation on sense perception and its objects, but rather as a kind of 'unavoidable, conceptual shadow of our immanent experience' (Bird, (2006) p.706). Thus conceived, noumena could stand to denote, in extremely broad terms, an inevitable conceptual residue left over from experience, whereas the thing in itself is conceived of within far-

more narrow parameters simply as a specific example of such, namely the thought of an object aside from our experience of it.

[t]he concept of a *noumenon*...does not signify a determinate cognition of any sort of thing, but rather only the thinking of something in general, in which I abstract from all form of sensible intuition.

(A252)

With this in mind, we can address the question of whether a distinction between noumena and phenomena could put right the concerns expressed over the inadequacy of the distinction between the thing in itself and phenomena. Indeed, the question at hand could be posed something like, granted that the thing in itself was found to be too narrow to accommodate the intelligible entities necessary for the possibility of morality, could intelligible entities themselves not be examples of noumena?

There are indeed important similarities between the intelligible entities and noumena that might speak in favour of such a reading, namely, both intelligible entities and noumena are conceived of as ‘given to the understanding alone and not to the senses at all’ (A257/B313). That is, the break with sensibility that noumena ushers in, and with which the thing in itself is so inextricably tied up (at least on the narrow reading of the thing in itself presently discussed) seems to place noumena and the necessary intelligible entities appropriately close. Building on this thought, we might also note that whilst the idea of the thing in itself (in this narrow sense) is an idea of an object *extrinsic* to our conscious experience, the idea of noumena is an idea of something *intrinsic* to our experience; the distinction is between, on the one hand, an extrapolation upon sensibility *and* the understanding (thing in itself) and on the other, an extrapolation of pure reason’s faculty of the understanding *alone* (noumena). In this sense then, the sort of inquiry that leads one to the conception of noumena looks a lot more like something akin to mathematical

exposition, and not anything like reflection upon sense perception¹². Thus the equation of intelligible entities with noumena *rather* than the thing in itself seems to fit more neatly with some of Kant's statements regarding the need to explain freedom and morality 'without going out of ourselves' (*C.Pract.R.*, p p.112/238-9)¹³. This is presumably how Kant sees the matter, at least at the moments when noumena and the intelligible are conflated.

Yet it is for familiar reasons that noumena simply fail to alleviate the particular tension highlighted in the preceding. Once again, critical ignorance is delivers the final blow, rather than offering deliverance. The concept of a noumenon, Kant tells us explicitly, is 'merely a **boundary concept**, in order to limit the pretension of sensibility, and [is] therefore only of negative use' (A255/B311). Yet if this is the case, then noumena falls prey to all the same pitfalls that faced the thing in itself. That is, thus conceived, noumena stands as a conceptual counterpart to spatiotemporal phenomena, since it has only negative use, and nothing positive can be said of it, such that it too is 'empty (for us)' (A255/B311). Yet if this is the case, then surely it also stands as a conceptual counterpart to our experiences of *meaning*, *worth*, *purpose* and *possibility* without which morality simply crumbles away. Precisely because it only acts *only* as a boundary concept, itself *empty*, and thus stands to counterpoise everything necessary for an experience of morality, the introduction of a distinction between noumena and phenomena fails to put right the relevant deficiencies of the distinction between the thing in itself and phenomena. I labour this point further if only because of its importance for what will follow. Neither the thing in itself nor noumena are such that we could possibly have knowledge of them. Thus, nothing positive pertains to our notion of either, and they each find merely negative use. The upshot is that our experiences of *meaning*, *worth* and *possibility* only find legitimate application within human experience – that is, they are all *relative* and *peculiar* to the human condition, and cannot be ascribed to the intrinsic nature of things (*ie* not to the thing in itself) and further, have no *general intrinsic*

¹² so far as the two can be distanced within the critical philosophy.

¹³ This theme and its bearing on the objectivity of the moral law will be examined in more detail in the next section.

application *per se* (ie find no place among noumena). Yet, Kant's wider moral project, with its invocation of a community of rational beings and the pursuit of a kingdom of ends requires that *meaning, worth* and *possibility* find precisely the kind of supra-human, if not *unconditional* application (ie, that beyond the merely contingent and limited human consciousness) that the restriction of noumena to a boundary concept with only negative use seems to rule out.

As a further thought, itself an extension of the last, we might also comment that if noumena are to be considered as the conceptual *shadows*, or discursive *residue* left over, indeed, resultant of, immanent experience, then noumena would have to *follow* experience. Yet Kant is clear – the moral law commands *immediately*. If intelligible entities *were* noumena, would they not come on the scene too late for them to be accommodated by the immediate nature of morality?

With these points in mind, we are left asking not only why a distinction between the thing in itself and phenomena might make possible the space for morality, but further, why we would think that the distinction between *noumena* and phenomena would be any more enlightening in this regard.

§2: *Kantian Constructivism*

In the first chapter we saw that within Kant's practical philosophy two distinct lines of argument can be discerned, one where the contrast between reason and spatiotemporal appearances makes the necessary space for intelligible entities, and another which introduces the thing in itself. Further it was noted that the thing in itself is ill-suited for Kant's practical philosophy since it fails to accommodate the intelligible entities. Given then that the thing in itself seems so antithetical to what Kant actually needs to get his moral project going it seems pertinent to ask whether if we ditch the thing itself we must also ditch the moral philosophy that Kant constructs on its back. That is, could we still do 'Kantian' ethics if we left behind Kant's own problematic appeal to the thing in itself?

On the face of it, the prospects seem encouraging. After all, we noted that the thing in itself is *superfluous* since Kant has the resources to resolve the Third Antinomy using the 'reason model'. Indeed, because Kant conflates the 'reason model' with the 'thing in itself model' so gratuitously, there is no shortage of textual accommodation for the centrality of reason right the way through Kant's ethics, from the most fundamental of topics, the Third Antinomy, through to the more familiarly 'ethical' topics, such as lying, suicide and the rest that are found in the explicitly moral works. Indeed, at times Kant himself is quite clear that whatever the account of the absolute character of the moral law may be, such an account must remain *immanent*, not transcendent (*C.Pract.R.* p.50/164, p.112/238). Might Kant's own use of reason in his practical philosophy then be a more promising candidate than the thing in itself for the basis of a broadly-conceived 'Kantian' ethics?

In this chapter I will look at three influential attempts to pursue just such a path by the ethical constructivists, John Rawls, Onora O'Neill and Christine Korsgaard. Ultimately I will suggest that these constructivist interpretations of Kant's ethics, though right to reject the thing in itself, neglect certain insights that Kant was keen to incorporate into his

own project about the way morality appears to us, and more broadly, that they consequently misunderstand Kant's conception of the task of moral philosophy.

§2:i) *Constructivism and Metaphysics*

'Constructivism', conceived broadly, will be taken to denote a paradigm for morals in which normative practical principles are ushered in by certain 'constructive procedures' of agents themselves. Though they are not always conceived of as incompatible with other paradigms of ethics where moral principles are taken by agents to be grounded in facts about the way things are *aside* from our experience them, constructivist models always stand in conscious distinction from them (see eg, Rawls (1975) p.286-8; (1993) p.116-8). That is, constructivist models tend to try and side-step many of the 'big' traditional metaphysical issues that have dogged moral philosophy, such as *eg*, offering comprehensive, (in Kant's terms 'transcendent') explanations or defences of moral values or the good life, and also, we should add, the ultimate metaphysical truth regarding human freedom.

...we do not require of a moral...judgment that the reasons for it show it to be related to an appropriate causal process... Rather, it is enough that the reasons offered be suitably strong. We explain our judgment, so far as we do, simply by going over the grounds for it: the explanation lies in the reasons we sincerely affirm. What more is there to say except to question our sincerity and reasonableness?

(Rawls (1993) p.118)

Note then that the kinds of concerns with which Kant wrestles in the Third Antinomy is quite foreign to the kinds of concerns of the constructivists. Historically problematic metaphysical questions are treated as off-limits for the constructivists, and the focus

instead is with the matter immediately at hand – namely, morality here and now. This is the sense in which for Rawls, moral theory is ‘independent’ from the rest of philosophy ((1975) p.286-302). Yet note also the similarity between the constructivists and Kant here. Though in the Third Antinomy and elsewhere, Kant wrestles with the historically problematic question of the metaphysics of human freedom, his ultimate position is one of ignorance. As we saw, Kant does *not* think that theoretical reason could ever offer a definitive answer to that question. Further, as the constructivists build their moral theory on the back of insulating it from metaphysics, it is on the back of this ignorance regarding freedom that Kant builds his ethics. In this sense then, though the parameters of their conversations are distinct, we might find enough similarity between them and Kant to think of the constructivists as broadly-Kantian.

§2:ii.i) *Rawls*

Since constructivism, thus conceived, more or less starts with John Rawls, and Rawls offers a constructivist reading of Kant’s moral philosophy, it seems only fitting to start with his account. Recall that Kant supposes that it is the appeal to the thing in itself that ensures the possibility of morality. Clearly for Rawls, by insulating moral theory from metaphysics (1975; 1993), such an appeal is ruled out. Further, as we have seen in chapter one, this seems fortuitous, since the thing in itself brings with it more trouble than good. More, Rawls agrees that the blind *is* of things apart from human reason cannot be the locus of morality. In his ‘Themes in Kant’s Moral Philosophy’ ((1989) hereafter ‘Themes’) Rawls comments that ‘there exists no moral order prior to and independent of the form of the procedure that specifies the content of first principles of right and justice among free and equal persons’ ((1989) p.97). Without the thing in itself, nor any other prior or ‘external’ set of moral principles are available to him, Rawls needs to find another candidate to open up the possibility of morality and he thinks he finds it in Kant’s conception of practical reason.

Rawls' basic idea is that we start from a normatively loaded, though uncontroversial, conception of persons as 'free and equal', and construct a set of principles that follow from this non-controversial account of agency. As such, moral commands gain their authority not from any fact to be discerned about the world aside from our experience of it, nor from any particularly complex moral psychology or hierarchy of values, but rather from a constructive procedure based on uncontroversial claims about persons. It is the *procedure* that specifies the first principles of right and justice (which we may here just call moral principles) and further, it is through that procedure that the *content* of the moral law is constructed. Note then that prior to the procedure there are simply no moral facts to be known. Rather, moral facts depend on the constructive actions of agents who go through a particular procedure.

Apart from the procedure of constructing the principles of justice, there are no moral facts. Whether certain facts are to be recognized as reasons of right or justice, or how much they are to count, can be ascertained only from within the constructive procedure, that is, from the undertakings of rational agents of construction when suitably represented as free and equal moral persons.

('Kantian Constructivism in Moral Theory' (1980), hereafter 'Kantian Constructivism', p.519)

Further, since the procedure is grounded firmly in practical reason, consensus in morality is accommodated.

...we must suppose, as Kant does, that whoever applies the CI-procedure, roughly the same judgements are reached, provided the procedure is applied intelligently and conscientiously, and against the background of roughly the same beliefs and

information. Reasonable and rational persons must recognize more or less the same reasons and give them more or less the same weight.

(‘Themes’, p.101, my emphasis).

To summarize Rawls’ ‘Kantian’ position, there are moral facts in the world, but only ones that arise via the constructed principles that follow from a mutual understanding of each other as free and equal persons. Further, objectivity of the judgements made according to such principles is ensured by the ‘background of roughly the same beliefs and information’ shared by agents. So whilst Rawls’ theory is constructivist in the sense that the foundational moral principles arise out of something *people do* rather than some fact about the world aside from experience, the procedure itself is not chosen, but is a consequence of the uniform character of the faculty of reason common to all, and the given social surroundings.

What justifies a conception of justice is not its being true to an order antecedent to and given to us, but its congruence with our deeper understanding of ourselves and our aspirations, and our realization that, given our history and the traditions embedded in our public life, it is the most reasonable doctrine for us.

(‘Kantian Constructivism’, p.519)

§2:ii.ii) O’Neill

Whilst O’Neill is broadly sympathetic Rawls’ project, she is surely right to point out the decidedly relativistic spin that Rawls, apparently inadvertently, adds to Kant’s thought in his account of the objectivity of the moral law. The problem, as O’Neill presciently notes, ultimately lies with Rawls’ starting point – that of an idealized conception of free and equal persons. Whilst Kant surely shares such an ideal conception with Rawls (though

retaining ultimate ignorance of the noumenal fact of the matter regarding freedom), it is *not* the starting point of Kant's *moral* philosophy. Rather, to the extent that Kant starts his moral philosophy with a conception of persons at all, he begins with the far more abstract conception of '*homo noumenon*', at other times with a 'subject whose actions can be *imputed* to him' (*MM*, p.50/223) or even more generally, with the concept of the unconditionally good will, taken in its broadest sense (*Gr*, 4:393), but *not* with socially embedded persons. Further, as O'Neill is at pains to point out, when Rawls lays out his reading of Kantian moral theory, it is not just any old notion of socially embedded persons that he begins with, nor simply with a general fact that people are socially embedded, but rather, with the specific conception of free and equal persons. It was commented above that such a conception was normatively loaded, though non-controversial. O'Neill's reply to Rawls here is that this starting point *is* controversial because it assumes that *just these* normative assumptions are generic or universal. As O'Neill sees the matter though, this is far from the case, with the result that Rawls seems markedly un-Kantian in his ensuing relativism.

Rawls [only offers a defence] of those deep principles of justice "we" would discover in drawing on "our" underlying conceptions of free and equal citizenship. This vindication of justice does not address others, who, unlike "us" do not start out with such ideals...[Rawls'] Kantian constructivism, it seems, claims only to offer a coherent articulation of the outlook of modern liberal societies. The Kantian ideal of the person is socially embedded, and antirelativism is not attainable. We are offered a coherent articulation of the deep moral commitments of "our" society.

(*Constructions of Reason*, (1989) hereafter *Constructions*, p.211)

O'Neill's alternative to Rawls' position is to move away from the socially constructed desires of persons, an admittedly heteronomous position for Kant (*Gr* 4:441-5), and

instead root the entire procedural account in rational agency. That is, O'Neill distances her account of how the scope of the moral law could be objective from Rawls' by abstracting from agents' desires, focusing instead on a minimal conception of agents' rationality and then considering what principles it would be possible for those agents to act on. O'Neil's constructivism might thus be seen as an important development on the 'reason model' identified in chapter one as central to Kant's moral thought, appealing to the minimal rationality of the individual to ensure the possibility of morality in a manner that avoids the un-Kantian implications of relativism found in Rawls' thought.

Everything...hinges on constructing principles of justice without presupposing a determinate ideal of the person...The core of any such construction is the thought that there are certain constraints on the principles of action that could be adopted by all of a plurality of potentially interacting agents of whom we assume only minimal rationality and indeterminate mutual independence.

(*Constructions*, p.215)

For O'Neill then, the relevant question is *not*, what desires may people have, nor which principles they might 'choose to live by', even less their contingent social position or surroundings, but rather the *modal* question 'what principles can a plurality of agents of minimal rationality and indeterminate capacities for mutual independence live by?' (*Constructions*, p.213). That is, with *desire* off the menu, O'Neill reconstructs the constructivist procedure such that the possible collective agency of a plurality of individual, yet interrelated, minimally rational beings plays the operative role in generating normative practical principles. The absolute character of morals would then follow from the notion that '[p]rinciples that cannot be acted on by all must be rejected by any plurality' (*Constructions*, p.215). Violence and coercion are offered as examples of the sort of things that the procedure would reject in the construction of principles, since 'principles of action that hinge on victimizing some, whether by destroying,

paralyzing or undercutting their capacities for agency for at least some time and in some ways, can be adopted by some but cannot be universally adopted' (*Constructions* p.215-6).

Yet, with desire off the menu, it is incumbent upon O'Neill to offer an account of the normative *pull* of moral principles. *How* exactly might this minimally rational agency generate normative authority? Whilst Rawls could fall back on an idealized notion of persons reciprocally respecting each others' freedom and equality, since O'Neill does not accept this as a legitimate *starting point*, she cannot rely on it and has to find another way of accounting for the normative force of practical principles. That is, if practical principles are the consequence of constructive procedures, and the principles are to have normative force, they must somehow attain that force *from* the procedures. O'Neill is committed to something like the claim that if the procedures are based on reason, and the principles that arise out of the procedures have a normative *pull* or *force*, then reason itself must have the required normative authority and force itself, such that it is imparted to the principles. To the further question of what justifies or confers upon reason this initial *pull* or *force*, we have already seen the answer. It is by virtue of the fact that the principles can be 'universally adopted', that is, it is by virtue of the *public* aspect of reason, that *all* can assent to it that reason gains its authority.

The most basic requirement for construction by any plurality of agents must then be... no more than the requirement that any fundamental principles for thought and action we deploy be ones that it is not impossible for all to follow... Here we begin to understand why Kant held that the Categorical Imperative was the supreme principle not just of practical but of all reasoning.

(*Constructions*, p.19-20).

Note that on the surface, O'Neill's account seems to look closer to Kant's than Rawls' at least at those moments when Kant is at pains to dissociate his moral philosophy from mere 'practical anthropology' (*Gr* 4:388), from 'heteronomy' in morals (*Gr* 4:441-5), from 'eudaimonic' grounds for morality (*Gr* 4:442-3, *C.Pract.R.*, p.94/217) indeed, inclination generally (*MM* p.44/216), and instead ground morality 'without going out of ourselves' but rather, by means of 'our own reason' (*C.Pract.R.* p.112/238-9, *see also*, *eg*, p.32/143). Any one of such charges could be conceivably be levied *against* Rawls, since Rawls not only places the empirical desires of agents centre stage, but further, grounds the normative *pull* of the moral law within the social setting in which we find ourselves – and as O'Neill points out, it is indeed the social setting in which “*we*” find ourselves that ends up doing the work for Rawls' Kant.

Further, if successful, O'Neill's account would steer clear of the sort of 'obscure and panicky' metaphysical claims (*Constructions*, p.210) which, as we saw, loom large over Kant's actual account. For O'Neill, the constructive procedure would only need to invoke the limited rationality of agents such as we encounter *within* experience. This non-metaphysical account would also appear to avoid the relativistic pitfalls of Rawls' account, since the appeal is only to a *minimal rationality*, such that is designed to exclude socially constructed desires, and get to the bare core of agency. Thus, if successful, O'Neill's account would avoid an excessive metaphysics, and also Rawls' un-Kantian relativism, whilst maintaining the 'immanence' of the moral law with which Kant is concerned, at least in *C.Pract.R.* (p.50/164, p.112/238).

It is important to note that there are tensions and questions intrinsic to O'Neill's account. First, it might be suggested that O'Neill's account of reason wrongly confuses the nature of reason with its consequences. Consider the following claim:

- 1) If something is rational, then anyone could assent to it.

This much seems uncontroversial. Yet note that O'Neill surely invokes an inverted form of this claim in her procedural constructivist account:

2) Rationality is *constituted* by the fact that anyone could assent to it.

If 1) can be admitted to be a legitimate claim as to the nature of reason, then of 2), it could be commented that it mistakes the consequences of rational action for what constitutes rationality itself. Indeed, if it is uncontroversial, as I think it is, to admit that the former is a more common sense view of reason, then O'Neill's adherence to 2) seems somewhat forced, if not confused. That is, O'Neill suggests that a proposition's rationality is *constructed and determined* as a consequence of a particular procedure, and we might doubt that claim - perhaps the procedure is doing too much work here.

Similarly, it might also be questioned whether O'Neill's constructivism actually can deliver the strong normative *force* it claims to accommodate. At issue here is not the rationality of principles *per se*, but rather, the link between that rationality and the necessary normative *pull* or *force* that moral principles have on us. This normative pull could be accounted for in one of two ways: by virtue of some external authority, say, a lawgiver such as God, or by some internal authority. In typical constructivist style, the possibility of an external authority is to be rejected as a candidate for this authority. Thus, O'Neill's constructivism looks to internal authority in an attempt to side step controversial metaphysical commitments. O'Neill's next claim, indeed, her 'big' claim is that the only internal authority that seems plausible is the public use of reason.

There is no lofty position above the debate, as perhaps there might be if human reason had a transcendent source. There is only the position of one who strives to reach and understand the perspectives of others, and to communicate with rather than past them

(*Constructions*, p.46-7).

Yet this is, by O'Neill's own admission, merely a negative argument. 'All that has been established for beings who share a world is that they cannot base this sharing on unadoptable reasons' (*Constructions* p. 27, *see also* p.20, 22). The question then arises, since moral force seems necessary for her picture, is it not still incumbent upon O'Neill to offer a *positive* account of *how* public reason could generate the requisite sense of

normativity? Further, this seems particularly pressing, since without such an account we run the risk of simply conflating reason with whatever may happen to be popular enough to provide some guiding *pull* within the public realm. Unless such a distinction can be made, it seems that just about any principle could be granted the status of a moral principle, so long as it was suitably popular and were agreed to in accord with the public use of reason. But this surely comes at the cost of losing Kant's distinction between hypothetical and categorical imperatives (*Gr* 4:414-5), indeed, even moral and non-moral principles (*Gr* 4:416), and that seems like precisely the sort of distinction that we want to keep, if we want to be faithful the way morality appears to us in experience. Kant is, I suggest, quite right to insist that the regardless of the metaphysics of the matter, there is just something *unique* about our moral experiences. As Kant puts it, 'the categorical imperative alone has the tenor of a practical law' (*Gr* 4:420 my emphasis). On Kant's account, the mere fact that people can collectively decide to follow a given principle does not necessary make it moral, even if they decide to it under the constraints of public reason, since to follow some such principles may be to use certain means to secure ends, rather than to preserve ends in themselves. The distinction could be drawn in terms of a presumably non-moral decision to follow a principle that will lead to *eg*, playing a game of football and a decidedly moral decision to follow a principle that will *eg*, save a life. It is entirely conceivable that each principle could be constructed in accordance with the public use of reason and that neither relies on any external authority, and yet surely only one is *moral*. It seems then that that which ensures the link between reason and what Kant thinks of *moral* force may not be accommodated by O'Neill's negative account of the link between public reason and normative force. That is, O'Neill's account of normative force could do with some re-enforcement if it is to meet Kant's awareness of the phenomenology of morality.

In short then, whilst O'Neill seems closer to Kant than Rawls in consciously avoiding relativism, and more, her notion of minimal agency has important resonances of the 'reason model' of morality attributed to Kant earlier, it might be thought that her account sees the constructive procedure doing too much work in creating or constituting the rationality of principles and propositions. More, her account of the normative pull of

morality seems to fall short of Kant's awareness of the privileged appearance of morality in experience.

§2:ii.iii) Korsgaard

Korsgaard too places great emphasis on the need to avoid 'ontological or metaphysical commitments' in moral philosophy (*Creating the Kingdom of Ends* (1996) hereafter, *Creating*, p.305n17). She shares with Rawls and O'Neill the view that value and normativity are not *out there* in the world to be intuited or discovered, but rather are generated and imposed upon the blind *is* of the world aside from experience by the exercise of rational agency. Korsgaard is clear - she takes this to be the only really legitimate position to hold, given the modern scientific worldview.

If the real and the good are no longer one, value must find its way into the world somehow. Form must be imposed on the world of matter....The ethics of autonomy is the only one consistent with the metaphysics of the modern world

(*Sources of Normativity* (1996), hereafter, *Sources*, p.5)

The stakes are high then, and Korsgaard's account is more stringently constructivist than either Rawls' or O'Neill's. Korsgaard takes the constructivist discussion deeper, expanding the themes at play from the now familiar issues of whether and how value and subjective ends are accommodated within the framework of the blind *is* conception of the world aside from our experience, and looks beyond, to a key question central to any account that lays claim to the "Kantian" heritage, namely, given our admittedly finite standpoint and the prohibition on 'metaphysical' appeals to any external authority, can we save 'absolute' moral claims?

Korsgaard's argument, roughly sketched, maintains that *qua* rational beings, we must have reasons for acting, that is, our practical activity is such that we act for reasons, not blindly. This model of rational agency itself presupposes that some actions are viewed as better than others, which in turn presupposes that we view some reasons as *genuinely* good reasons for acting. This requires regarding some reasons as corresponding to some genuinely good ends. Yet, to regard some ends as genuinely good, given that external authority is ruled out by the modern scientific worldview, requires that 'we must regard ourselves as capable of conferring value upon the objects of our choice, the ends that we set' (*Creating* p.260). The next stage of the argument maintains that it is only by virtue of regarding ourselves as having unconditional, or absolute value that we could conceive of transferring or conferring this value onto objects of our choice. That is, we must regard ourselves as capable of ensuring the judgement that something is genuinely good, and that this requires regarding ourselves as having an absolute value. Thus, in order to coherently conceive of our value-laden experience, we must regard ourselves as unconditionally valuable. Note then, that we must regard ourselves in this way *because of our rational nature*. Yet, if we regard ourselves as unconditionally valuable because of our rational nature, then we must also be committed to regarding other beings with an equally rational nature as possessing an equally unconditional value. Thus, our rational agency, that is, our practical activity in the world, entails a commitment to regarding all rational agents as unconditional ends in themselves. So on Korsgaard's reading, we value rational agents unconditionally then not because of a fact *about them* that we intuit or recognize, but instead as a consequence of the commitments tacit in our conception of *our own agency*.

This line of thought leads Korsgaard to the somewhat unexpected, indeed counter-intuitive claim that the value of other people lies exclusively in something that I do. Thus Watkins and Fitzpatrick (2002) are right to raise Blackburn's quasi-realist account of what we ordinarily think is wrong with slavery against Korsgaard. The thought is that it seems far more in line with our regular way of thinking about morality to say that slavery is wrong *not* because of any action of mine (assuming that I am not involved in the slave trade), but because there is just something straightforwardly wrong about *another person* having to experience something like that. To explain the wrongness of slavery by virtue

of some feature about myself seems to take a needless and distracting detour down the wrong path. Morality just does not seem that way to us in experience.

To cash out the wrongness of such an action and its normative force for me in such a way that requires a detour through a story about what I have to do in order to exercise my will at all seems like a move precisely in the wrong direction. It does not seem true to ordinary moral experience, which certainly does not represent other people's value and significance for us deriving from commitments bound up with the exercise of our own wills under certain generic constraints inherent in the nature of willing.

(Watkins and Fitzpatrick (2002), p.361)

We might say then that like O'Neill, Korsgaard is faithful to the reason-heavy aspect of Kant's moral thought, that she perhaps does a better job of accounting for the unique, 'absolute' character of morality in experience, and more, that she is surely right to do all this while leaving behind any appeal to thing in itself. Like O'Neill though, Korsgaard fails to capture something quite fundamental about the phenomenology of morality. O'Neill's constructivism seems to mistake the rationality of principles *for* their constructed-ness, and also falls short of accommodating the strong normative force unique to the way morality strikes us. Whilst Korsgaard does offer an account of precisely this normative pull in her account of the uniquely absolute character of moral commands, she does so at the cost of some key commonsense intuitions about morality. This is particularly prescient in light of the question of staying faithful to a broadly-conceived 'Kantian' ethics, since Kant himself is at pains to maintain our commonsense thoughts about morality. It is only thus that he can proceed 'from common rational cognition to philosophy' (*Gr* 4:413). As Kant asks so presciently asks, given the importance, and he thinks *correctness*, of our common appreciation of morality in experience:

Would it not therefore be more advisable in moral matters to leave the judgement of common reason as it is and, at most call in philosophy only to present the system of morals all the more completely and apprehensibly and to present its rules in a form more convenient for use (still more for disputation) but not to lead common human understanding away from its fortunate simplicity and to put it, by means of philosophy, on a new path of investigation and instruction?

(*Gr* 4:404 my emphasis)

§2:ii.i) *Kant and Immediacy*

Central to each of the constructivist accounts is the importance of the subject's deliberation, private or public, in moral actions and rational agency. The key idea that unites all the constructivist readings is that reason is the only reliable guarantor of the right action, and so the right thing to do must be worked out and consciously assented to by the subject. As we saw, for Rawls, this meant that moral principles and content could only be established through a specific constructivist procedure, for O'Neill, that such a procedure was rooted in the public use of reason, and for Korsgaard, that it is only by reflection on our rational agency that the unconditional value of others is coherent, given the modern scientific worldview. More, it was noted that this emphasis on reason as the basis of morality seemed to resonate with certain aspects of Kant's own moral philosophy, namely the 'reason model' of morality discussed above.

Yet in *Gr*, and in *C.Pract.R* especially, we are repeatedly reminded of the immediate nature of the moral law's command over the will. This of course raises the question of why we should think that a constructivist picture requiring elaborate procedures and forms of deliberation could really be compatible with Kant's actual moral thought. As we have already mentioned, Kant has a tendency to view common moral thought as broadly in the right place already. In the preface to *Gr*, Kant is quite clear that really, offering a philosophical account of the nature of morality is not as pressing a concern as a philosophical account of theoretical reason, since on the whole, people are pretty close to, and easily brought in line with, moral action.

Indeed there is really no other foundation for a metaphysics of morals that the critique of *a pure practical reason*, just as that of metaphysics is the critique of pure speculative reason, already published. But in the first place the former is not of such utmost necessity as the latter because in moral matters human reason can easily be brought to a high degree of correctness and accomplishment, even in the most common understanding

(*Gr*, 4:391)

Note then that Kant seems to have a very different conception to the constructivists of the ‘state of play’ regarding moral philosophy. Kant tends to think that on the whole people just are more or less moral, and that their pre-philosophical thoughts and beliefs about morality are fine left as they are. It seems then that in contrast to the constructivists, Kant is not really concerned with ‘constructing’ moral norms at all. Rather, he views pre-philosophical moral opinion as broadly in the right place, and the sort of problems he engages with arise specifically within the confines of philosophical investigation. That is, the philosophical question of the ultimate *foundation* of morality implies, for Kant, no revision of moral norms themselves, nor any revision of how they seem to us in experience, and I suggest, no prescriptive cannon of morally acceptable or prohibited acts. In contrast, procedural constructivists tend to start out from a position of general disagreement regarding the morally good life, and then literally ‘construct’ moral norms that provide a framework within which we can approve or condemn specific actions with the suitably justified *moral authority* of ‘right’ and ‘good’. For these constructivists then, it is precisely because the morally good act is not already clear to us from the get-go that we require some procedure. This means that only once we construct the procedure can we construct the parameters of moral goodness. Whereas Kant’s moral philosophy begins with the assumption that on the whole, people more or less already know what to do and how they ought to act, it is because they start from a position of ignorance regarding precisely these questions that procedural constructivisms like Rawls’ and O’Neill’s arise at all.

Further, I take it that it is by virtue of his acceptance of the *immediacy* of the moral law's command over us that Kant thinks people more or less already know what to do.

What I cognize immediately as a law for me I cognize with respect, which signifies merely consciousness of the *subordination* of my will to a law without the mediation of other influences on my sense. Immediate determination of my will by means of the law and consciousness of this is called respect, so that this is regarded as an *effect* of the law on the subject, and not as the *cause* of the law.

(*Gr*, 4:401n)

The implication of this passage is clearly in line with the stated aim of *Gr*, namely to identify the supreme principle of morality. What is not so clear from this passage though is where exactly anything like a constructivist picture of *procedure* fits into an explanation of moral action, such as was the concern of Rawls and O'Neill. Rather, for Kant, in moral action the will just is immediately determined by the moral law (hence the feeling of 'respect') and moral philosophy's scope looks limited to more abstract questions regarding the limit and constitution of reason, theoretical and practical, an example of which is to settle to what extent form or content determines the will when its action is moral. Indeed, I suggest it is precisely *this* kind of thought about the *descriptive* role of moral philosophy that motivates Kant's insistence at *Gr* 4:404 that in moral matters philosophy should leave things as they appear to the 'fortunate simplicity' of common moral cognition. The procedural constructivist claim that moral action *only* arises out the correct type of procedure thus seems quite foreign to Kant's thought.

The immediate determination of the will by the pure form of the moral law is also a prominent, repeated theme within *C.Pract.R.* spanning most of the work.

Reason, with its practical law, determines the will immediately

(*C.Pract.R.* p.24/133)

The practical rule is...unconditional, and hence it is conceived *a priori* as a categorically practical proposition by which the will is objectively determined absolutely and immediately (by the practical rule itself, which thus is in this case a law); for *pure reason practical of itself* is here directly legislative.

(*C.Pract.R.* p.31/142)

It is a sublime thing in human nature to be determined to actions immediately by a purely rational law;

(*C.Pract.R.* p.125/255)

...only when ...respect has become active and dominant does it allow us by means of it a prospect into the world of the supersensible, and then only with weak glances; all this being so, there is room for true moral disposition, immediately devoted to the law

(*C.Pract.R.* p.158/294-5)

The centrality of this theme within *C.Pract.R.*, namely that of the immediate character of the moral in human life, is likely related the reconceptualization of the objective reality of the moral law and freedom into a ‘fact’ of experience. Whilst this may go some way to explain the prevalence of the theme within Kant’s exegesis, it fails to help reconcile the notion with the procedural constructivist commentators. Such commentators insist that the conscious deliberation of rational agents *actually generates* the specifically *moral* content of principles. Yet on Kant’s conception of morality, agents are immediately determined by reason and philosophers arrive late to enquire abstractly as to the supreme principles underlying such an occurrence. Indeed, this aspect of Kant’s thought seems to imply a serious reconsideration of the ‘Kantian’ roots of constructivism. As Allen Wood puts it:

If we are to understand properly the spirit of Kantian ethics...we must learn to ask far less of a “supreme principle of morality” than moral philosophers often ask. The function of a supreme principle of morality, then, is *not* to tell us directly,

from day to day and minute to minute, through some canonical process of moral reasoning to be applied exactly the same way to all situations, exactly which actions we should (and should not) be performing and precisely how we should be spending our time... The correct interpretation of Kant's formulation of the supreme principle of morality exhibits the principle as concerned not with devising a "CI" decision procedure, or a ravishingly subtle and clever calculus but rather with stating the ultimate value on which moral rules and duties may be grounded.

(Wood (2008) p.305-6).

On such an interpretation then Kant's moral philosophy seems quite distant from the project of 'constructing' a set of moral norms, and far to closer to Schopenhauer's dictum that it would 'be just as foolish to expect our moral systems and ethics would create virtuous, noble people, and holy men, as that our aesthetics would produce poets, painters, and musicians', since 'philosophy can never do more than interpret and explain what is present at hand' (Schopenhauer (1958) p.271). With this in mind, it becomes clear just why Kant is so keen to include the notion of immediacy and, we might add, the unique, privileged *feel* of morality and its pull that proves so troublesome for O'Neill and Korsgaard¹⁴. Kant's aim is not to design an infallible procedure that generates the moral character of certain actions, still less to derive the definitive canon of 'morally approvable acts'. Rather, he takes moral phenomenology the way it is, immediate and unique, and works back *from* there to a derivation of the processes and fundamental rationality that must be in place. It is, I suggest, in *this* sense that we ought to understand Kantian ethics as fundamentally rooted in reason – the sense in which we examine the underlying rationality supporting our moral experience.

The more procedural constructivisms of Rawls and O'Neill thus look significantly further away from Kant than Korsgaard, since her constructivism is rooted more squarely in a derivation of the principles underlying rational agency than in the importance of a

¹⁴ Note, I am *not* here committing Kant to the position that morality is reducible to a feeling. Rather I am suggesting that Kant is well aware that morality has a unique feel about it when we encounter it in experience.

procedure for discerning which principles are properly moral. In this sense then Korsgaard may be thought the more legitimate heir to a broadly-conceived ‘Kantian’ ethical project rooted in an examination of human reason. Yet as we saw, Korsgaard too seems to miss something quite fundamental about the way morality appears to us. As the slavery example brings out so clearly, in the sense that her account is quite so obviously at odds with how we ordinarily think about morality, Korsgaard’s ordering of priorities within her moral theory seems decidedly un-Kantian. We may say then that Korsgaard too reveals her distance from Kant insofar as Kant himself is, as we have seen, concerned first and foremost to treat morality as we find it, and then to derive the principles that underlie it in such a way as to leave everything where it was before the inquiry began, such that the ‘fortunate simplicity’ of our moral experience remains undisturbed.

§2.ii.ii) *Kant and Conscience*

I have suggested that Kant’s moral philosophy is properly viewed as an analysis or derivation of morality, and one that strives to keep the *feel* of morality in experience intact, indeed untouched. The project of the ‘construction’ of moral content, or a position of moral authority, a space from which to praise or condemn with *moral* force thus looks quite distinct from Kant’s own position. Constructivists will of course object, suggesting that their project finds ample textual support from Kant’s moral works. Indeed, referencing *Gr*, Rawls reads Kant as providing a four step ‘categorical imperative procedure’ by which maxims can be ‘tested’ for their moral credentials, such that ‘[u]nless a maxim passes the test of that procedure, acting from that maxim is forbidden’ (Themes p.82-7). O’Neill also quotes from *Gr* when she writes that ‘[t]he best-known version of [Kant’s] procedure of construction, the C[ategorical] I[mperative], is the formula of universal law, which enjoins agents to “act only in accordance with that maxim through which you can at the same time will that it become a universal law”: agents should reject principles of action which (they take it) cannot be adopted by all.’ (Freeman (*ed*) (2003) p.355). Once again Korsgaard seems closer to Kant than Rawls or O’Neill in correctly describing Kant’s project in *Gr* is ‘a motivational analysis’ (p.66). Yet she too reads *Gr* in the same prescriptive light as Rawls and O’Neill, suggesting that

‘Kant’s achievement is to argue from [a] feature of right actions to a substantive moral principle which identifies which actions are right’ (p.55) via a ‘universalizability test’ (p.66-72). This sort of reading has led to a caricature of Kant’s moral philosophy in which agents are urged to pause, think about whether the maxim of their proposed act is universalizable, and then either proceed or abstain according to the result of that ‘test’. Worse still, that universalization somehow *makes* moral content. More, I suggest that it is precisely on the back of this caricature of what Kant’s moral philosophy actually amounts to that constructivists regard themselves as ‘Kantian’ – it is in this procedural ‘universalizability test’ that constructivists see their ‘Kantian’ roots. Given that I have suggested that this is *not* the conception of moral philosophy operative in Kant’s work, some account is required to put the constructivists right – Kant’s references to the formula of universal law need explaining in a way that frees him from this prescriptive baggage.

I have suggested that a better position from which to read Kant’s moral philosophy is to understand him as starting with people’s pre-philosophical experience of morality and then analysing or deriving the underlying commitments that must be in place, given its unique, immediate character. On this conception the questions that Kant is concerned with arise not because moral norms need to be ‘constructed’ by a ‘test’ or ‘procedure’. Rather, Kant views agents as generally already knowing through common moral cognition what the moral law asks of them and they know this prior to any ‘universalization test’. I suggest that Geiger’s understanding of the role of universalization within Kant’s moral thought as articulating ‘the shape of ordinary moral self-criticism’ (Geiger (2010) p.286) is particularly insightful here. Geiger’s claim is that Kant’s use of ‘universalization’ is ‘heuristic’ ((2010) p.286). Universalization serves to reinforce our common, pre-philosophical understanding of morality by bringing to the fore commitments that might otherwise be overridden by non-moral concerns such as crude self-interest. Universalization then serves to ‘bring an idea of reason closer to intuition (by a certain analogy) and thereby to feeling’ (*Gr* 4:436) by emphasizing and clarifying our pre-philosophical moral commitments. As Geiger puts it: ‘[b]y *thus making the transgression perfectly vivid* [universalization] *provides a heuristic tool for moral*

self-criticism...What the examples show is not the generation of new knowledge, but implicit knowledge made explicit...In this lies their heuristic value.’ ((2010) p.286). Universalization then offers a philosophical tool with which we can more fully understand the implied commitments of our pre-philosophical moral opinions. What universalization does *not* offer is the creation of moral content, still less a privileged position from which to judge and condemn. This reading finds ample textual accommodation, fitting neatly with Kant’s concern, noted above, to faithfully accommodate, and *not* corrupt, morality as we find it in experience - unique and immediate, not to mention his stated overall conception of the task of the moral philosopher as analysis or derivation of the fact of morality in experience (*Gr* 4:404-5; 4:409), and his qualifications regarding the merely *subjective* necessity of universalization (‘Thus people say: If *everyone* permitted himself to deceive when he thought it to his advantage...’ (*C, Pract. R.* p.73/193 (my emphasis); see also *Gr* 4:412-4; 436-7).

With universalization thus conceived as giving philosophical voice to pre-philosophical moral commitments, Kant’s moral philosophy seems quite far from the prescriptive reading accorded to him by the constructivists. Insofar as Kant’s ethics is prescriptive at all, his recommendation seems to be that we ought to carefully examine our own pre-philosophical commitments in order that we can recognize morality more clearly. In short, Kant’s moral project, thus conceived, amounts to a rigorous philosophical investigation into our common, immediate, pre-philosophical moral voice. It is then interesting to note that Kant’s mature thoughts on morality turn to the authority of *conscience*.

In *MM* Kant tells us that ‘every man, as a moral being, *has* a conscience within him originally’ (*MM* p.202/400). Conscience, he tells us, ‘is an unavoidable fact’ (*MM* p.202/400). It is simply part of our moral phenomenology, described as ‘practical reason holding man’s duty before him for his acquittal or condemnation in every case that comes under a law’ (*MM* p.202/400), and later using his much-favoured legal analogies,

compared to ‘an *internal court* in man (“before which his thoughts accuse or excuse one other”)’ (*MM* p.233/438) and ‘the inner judge of all actions’ (*MM* p.234/439).

Crucially, Kant accords conscience incredibly high status, commenting that ‘if someone is aware that he has acted in accordance with his conscience, then as far as guilt or innocence is concerned nothing more can be required of him’ (*MM* p.202/401). It seems then that quite the contrary to the claims of constructivists that Kant wants us to subject maxims to a ‘universalization test’, in fact, insofar as moral praise and blame are concerned, Kant seems content to ask simply whether an act was performed *in good conscience*. That is, conscience as ‘internal court’, the judge of all actions, makes no reference to universalization. Rather, the analogy of the court brings to the fore a more simple picture of accountability and responsibility to one’s own pre-philosophical conscience, and *not* the formulaic moral algebra the constructivists propose. More, there is an awareness of ignorance and fallibility at play in Kant’s thought that stands to distance him yet further from the constructivist emphasis on universalization.

Kant is clear that ‘an *erring* conscience is an absurdity’ (*MM* p.202/400). Kant’s thought here seems to be that conscience cannot err in the sense that it might fail to pass judgement one way or the other. As Kant puts it, conscience ‘is not something that he himself (voluntarily) *makes*, but something incorporated into his being. It follows him like his shadow when he plans to escape’ (*MM* p.233/438). Yet though conscience cannot *but* pass judgement on what is and is not a duty, Kant admits that ‘I can indeed be mistaken at times in my objective judgement as to whether something is a duty or not’ (*MM* p.202/401). So conscience *can* err in the sense that it can mistake non-duties for duties. What it cannot err in is judging *that* something is a duty or not. Note though, that for Kant it seems that whether the maxim we are following *is actually* a duty is secondary in terms of moral praise and blame to whether or not the action is performed in conformity with conscience. Thus one ‘cannot be mistaken in my subjective judgement as to whether [they] have submitted it to my practical reason (here in its role as judge) for such a judgement’ (*MM* p.202/401) and ‘as far as guilt or innocence is concerned nothing more can be required of [agents]’ (*MM* p.202/401). Kant thus seems committed to the

position that ignorance and fallibility regarding the *objective* question of what my duties actually are do not bear on the moral praise or blame I deserve. Instead, since ‘nothing more can be required’ of an agent than ‘that he has acted in accordance with his conscience’ (*MM* p.202/401) it is ok, Kant thinks, to be wrong about the ultimate moral fact of the matter, so long as the action is assented to *in good conscience*. Far from devising a generative universalizing calculus for morality by which we judge the moral correctness of maxims, Kant seems committed to a far leaner prescriptive position than the constructivists maintain.

Perhaps a little surprisingly, the prescriptive aspect of Kant’s mature ethics does not consist in the assertion of a *duty* to follow one’s conscience (*MM* p.202/401). Recall that conscience speaks to us so persistently, whether we like it or not, indeed even when we try to escape it, and more that conscience can get it wrong on the crucial question of discerning duties from non-duties. In this context it is perhaps understandable that rather than there being a duty to follow one’s conscience, instead there is Kant thinks, an undeniable obligation to take steps to inform ourselves and our conscience as best we can to the nature and character of our duties, in the hope that we get it wrong a little less.

It is incumbent on him only to enlighten his *understanding* in the matter of what is and is not a duty... The duty here is only to cultivate one’s conscience, to sharpen one’s attentiveness to the voice of the inner judge and to use every means to obtain a hearing for it
(*MM* p.202/401)

This, I suggest, is for Kant, the true ‘state of play’ for moral philosophy. On the whole people do not need much in the way of moral education – common pre-philosophical conscience has a ‘fortunate simplicity’ in its ability to guide us correctly. Nonetheless, experience is constituted by a plurality of voices and the heteronomous voice of the non-moral can sometimes be mistaken for the autonomous voice of morality. To the extent that philosophy can *help* in moral matters, it can do so simply by investigating the distinct, indeed unique and immediate *feel* of, and perhaps also bring out some of the

implicit logic within that pre-philosophical ability to get it right, in the hope that perhaps our fallible conscience might err a little less. The constructivists mistake the complexity and sophistication of Kant's analysis of our duties for a comprehensive, prescriptive project. In fact, Kant's moral project is far more modest – he wants to accurately *describe* morality, so that we might be able to discern it a little better in the troublesome cases where conscience mistakes non-duties for duties.

Ultimately though, Critical ignorance commits Kant to appreciating the wider irony of his moral project. At the most fundamental level, morality for Kant must always remain a mystery due to his unfortunate commitment to the unknowability of transcendental idealism's thing in itself, and the examination of conscience is no exception. As Kant reminds us in an important footnote to a discussion of conscience:

man as the subject of the moral lawgiving which proceeds from the concept of freedom and in which he is subject to a law that he gives himself (*homo noumenon*) is to be regarded as another (*specie diversus*) from man as sensible being endowed with reason, though only in a practical aspect – for there is no theory about the causal relation of the intelligible to the sensible... Our reason cannot pursue further his power...in this function; we can only revere his unconditional *iubeo* or *veto*
(*MM* p.234n/439n).

Ultimately, how conscience operates and the details of why it judges the way it does are unknowable. The best we can hope for from our enquiries into moral philosophy is that by some cosmic justice, the larger cosmic *injustice* of ignorance might, perhaps, be put right. But Kant must surely be committed to only ever viewing this possibility with, at best, a certain irony if he is to remain true to his Critical ignorance. At the most fundamental level, the workings of conscience must remain, for Kant, a mystery, and our efforts to enlighten it must remain, at best, a stab in the dark. Once again though, we notice that precisely that which Kant needs to get his moral project off the ground is modally distinct from the perspective of the thing in itself – namely, precisely that sense

of meaning which is unique to the human standpoint. Insofar as Kant's analysis of conscience falls back on the familiarly problematic model of man *qua* unknowable noumenon mysteriously causing man *qua* phenomenon, it is unsatisfactory for all the reasons examined in chapter 1, and, as we saw, no amount of ignorance can put that right.

§3: *Fichte's Ethics*

To recap, in chapter 1 it was suggested that the thing in itself, since modally distinct from intelligible entities, is ill-suited to Kant's practical philosophy. With its inappropriateness for the task in mind, in chapter 2 some recent influential 'constructivist' attempts to form a broadly-conceived 'Kantian' ethics without the thing in itself were examined. It was there suggested that in ignoring Kant's own concern to stay faithful to the phenomenology of our moral experiences, the constructivists misconceived Kant's moral project. Rather than offering a generative principle of moral content, still less, constructing a position of moral authority, Kant's moral project, it was suggested, is largely descriptive, though a prescriptive aspect was found in his discussion of the duty to 'enlighten' one's conscience. An implicit irony in Kant's ethics was also noted. Since the ultimate fact of how and why conscience operates the way it does is a *noumenal* matter, we must remain open to the possibility that our efforts to reduce the misidentification of duty are ultimately futile. More, since Kant's discussion of conscience falls back on his ill-conceived phenomena, thing in itself distinction, his account is problematic, since, as was suggested in chapter 1, the thing in itself is ill-suited to accommodating the intelligible entities necessary for morality – conscience included.

In this final chapter I turn to Fichte's ethics. Fichte's thought is interesting in light of the preceding since he attempts not only to accommodate freedom *without* appealing to the thing in itself, he also strives to preserve much of the moral phenomenology central to Kant's thought, but overlooked by the constructivists. Thus, Fichte's ethics may be read as a critique of Kant's insofar as his project meets Kant's aims without invoking the thing in itself, yet closer to Kant than constructivism insofar as he remains true to our experience of morality and our pre-philosophical opinions.

§3:i) *A non-thought*

Famously, Fichte rejects the thing in itself from his philosophy. Taken in isolation and at face value such a claim seems not simply bold, but appears to endorse a hopelessly ambitious metaphysical idealism which not only doubts, but rejects the existence of any thing external to one's consciousness and in the past Fichte has been thought of as committed to the position that the mind somehow literally makes reality¹⁵. But this reading has been widely disputed more recently. Martin suggests that Fichte's 'exclusion of things in themselves is methodological rather than ontological' (Martin (1997) p.66) and on the back of such a move, Ameriks suggests that the rejection of the thing in itself is 'consistent with taking Fichte to have a robust belief in physical reality, very much as we ordinarily think of it' ((2000) p.203). Sidestepping this somewhat vexed question of what ontology Fichte's rejection of the thing in itself entails, for our purposes it suffices to say that regardless of Fichte's ultimate ontological commitments, he is certainly committed to a blanket-ban on employing the thing in itself even in any hypothetically-positive explanatory capacity. So, to take a pertinent example, Fichte agrees with Kant that freedom, and more, the distinct character of the intelligible, needs to be accounted for if morality is defended. Yet, given the blanket ban on the invocation of even any hypothetically-positive explanatory capacity, he cannot even say, as Kant does, that *if* freedom is possible, it is only possible from the viewpoint of the thing in itself. If Fichte is to defend freedom and the intelligible entities by his own strictures, the defence must remain entirely immanent. Anything else would be, by his own account, 'a piece of whimsy, a pipe dream, a non-thought' (Fichte, 1988, p.71/I, 17).

Fichte's reasons for excluding any positive explanatory potential for the thing in itself are varied. At times he suggests that the very idea of a thing in itself is a contradiction in terms – that it is 'downright impossible' to even *think* of a thing in itself ((1988, p.73, I,19). At other times, he invokes Cartesian skepticism suggesting 'that we can never encounter anything independent of us, since everything is necessarily related to our

¹⁵ See, eg Coplestone, F. (1965) *From the Post Kantian Idealists to Marx, Kierkegaard, and Nietzsche*. Vol. 7 of *A History of Philosophy*

thinking' (*Introductions* p.86/501), at other times he refers to as 'a mere thought' (*Introductions* p.67-72/482-486) and 'something produced by our own thinking' (*Introductions* p.69/482). Clearly these different ways of referring to the thing in itself do not sit well together¹⁶ and recent commentators have paid most attention to Fichte's criticism of the notion of our being 'affected' by the thing in itself. Since this sits closest to the criticism of the thing in itself outlined in chapter 1, it will be our focus. Fichte's position here is that by Kant's own admission, 'the categories only apply only to what exists for us' (*Introductions* p.67/482), and so to assume that causality, itself only made possible through the categories could extend beyond our experience and somehow link us up with something that exists not simply *for us* but in itself, is to misunderstand the parameters of causality within the Critical philosophy. As Fichte writes of those 'Kantians' who assume this unjustifiable causal connection: 'Their thing in itself...is supposed to *have an effect* upon the I! Have they forgotten what they themselves just said [?]' (*Introductions* p.69/483). Fichte is acutely aware of the impossibility of causality applying beyond the scope of the categories, and so the very notion of a causally efficacious thing in itself is an unacceptable 'combination of...dogmatism...with the most resolute idealism' (*Introductions* p.69/483). Whilst this claim is admittedly distinct from that of chapter 1, which relied on a similar distinction, but focused more generally on the modal distinction between the sphere of human meaning and the thing in itself, the thoughts are very close to each other. What Fichte rules out is any sense in which the cause of our actions could, by Kant's own system, conceivably lie in the thing in itself. His point is, I take it, that 'thing' only finds proper reference within human consciousness – that 'thing' requires the entire Kantian cognitive artifice. Within Fichte's philosophy then, 'there is no talk whatsoever of any being, as being in itself; nor can there ever be any such talk...[since there is] being only for an intellect' (*Sittenlehre* p.23/IV, 18). So Fichte must agree, given his position on causality, 'being' and the rest, that the viewpoint of the thing in itself is modally distinct from the locus of the intelligible, even if his discussion does not flesh out this claim in terms of *meaning*.

¹⁶ As Martin has noted, the second type of claim seems to require a coherent, non-contradictory conception of the thing in itself 'for a negative but substantive purpose: to state...epistemological implications...' (Martin (1997), p.75-7).

§3:i.ii.i) *Freedom without the thing in itself*

Like Kant, Fichte's stated conception of what moral philosophy should aim for is, broadly, a description of the appearance of morality within experience. As Fichte puts it, 'ethics is the *theory of our consciousness* of our moral nature in general and of our specific duties in particular' (*System of Ethics* hereafter *Sittenlehre* p.21/ IV, 15). That is, like Kant, Fichte does not think that morality needs to be argued for, or shown to exist – they both take the appearance of morality in experience as their starting point and work back from there to a derivation of its underlying factors. Also like Kant, Fichte is clear that insofar as the deduction is successful, it must leave morality the same as before the inquiry began.

Just as one does not posit objects differently in space and time after one has obtained insight into the grounds of this operation... so does morality not manifest itself any differently in human beings after its deduction than before.

(*Sittenlehre* p.21/IV, 15)

More, like Kant, Fichte agrees that '[f]reedom is the absolute condition of all morality, and without it no morality whatsoever is possible' (*Sittenlehre* p.221/IV, 232). Yet unlike Kant, as we saw above, Fichte cannot appeal positively to the thing in itself in his philosophy, and so unlike Kant, he must derive freedom without positively invoking the explanatory potential of the thing in itself.

As we saw in chapter 1, in his moral philosophy, Kant often assumes that reason and the thing in itself are un-problematically interchangeable. That is, Kant indiscriminately (and unjustifiably) alternates between invoking a distinction between spatiotemporal appearances and the thing in itself on the one hand, and a distinction between spatiotemporal appearances and reason on the other. Speaking very broadly, like the constructivists (though in a thoroughly different tone and manner), having ditched the thing in itself, Fichte picks up on this second contrast, the 'reason model', distinguishing

the inert content of sensibility from the active, ‘self-sufficiency’ of reason, accommodating freedom in the latter.

Fichte thinks we are conscious of our freedom through ‘intellectual intuition’ (*Sittenlehre* p.60/IV, 59). It is important to note that Fichte’s sense of ‘intellectual intuition’ is distinct from Kant’s. As we saw above, Kant uses the term ‘intellectual intuition’ in *C.ofJ.* §76 when discussing the (for us impossible) gaze which intuits aside from the categorial cognitive artifice, and thereby also stripped of meaning, intelligibility and the rest. This is *not* how Fichte uses the term. Rather, for Fichte, intellectual intuition designates the form of consciousness’ self-reverting activity which Fichte thinks lies at the heart of all (properly-conceived) philosophy, is also the original form of both consciousness and the moral law, and demonstrates our freedom.

In thinking about the thinking of a wall, Fichte suggests, the philosopher ascertains the division of consciousness into subject and object. In our pre-philosophical thinking of the wall as object, there is an *unrecognized* activity, namely, that of the subject doing the thinking, as *distinct* from the object being thought, the wall. That is, ‘the object that is being thought is not supposed to be the thinker itself’ (*Sittnelehre* p.24/IV, 19). The first job of philosophy then is to make this implicit distinction *explicit*, and bring to the fore the otherwise ‘lost’ activity of thinking itself. As Fichte puts it:

A rational being ... loses itself, as subject, in the object. In philosophy however, everything depends on becoming acquainted with the subject as such
(*Sittenlehre* p.36/IV,31)

This first form of consciousness in which subject and object are taken to be entirely distinct entities is then contrasted with the form of thinking about oneself. In this latter case, whilst the thinker is taken as ‘object’, in the sense that the thinker becomes the object of thought, here, the object being thought is supposed to be *precisely* the same as the thinker doing the thinking. That is, in thinking about the thinker, an admittedly artificial abstraction of the philosopher, one obtains a *new* concept, one entirely

unobtainable through the thinking of walls, namely, the concept of a fundamental unity of subject and object. Self-reverting cognitive activity thus yields *new* content. More, this new content is *wrought of itself* – in thinking about itself as both subject and object, consciousness *provides itself* with an otherwise un-encountered (and Fichte proposes that it is otherwise un-encounterable) concept of the fundamental unity of subject and object. This new concept Fichte calls the concept of I:

when, in an act of thinking, the thinker and what is thought are taken to be the same and, vice versa, what arises in such an act of thinking is the concept of the I.
(*Sittenlehre* p.24/IV, 18-9)

Recall Kant's definition of freedom in the Third Antinomy: 'the faculty of beginning a state **from itself**, the causality of which does not in turn stand under another cause determining it in time in accordance with a law of nature' (A533/B561). Fichte is aware that whilst this philosopher's abstraction shows, at best, that this act of self-reverting consciousness demonstrates the beginning of a new state 'from itself', what he has not yet shown is that consciousness itself 'does not in turn stand under another cause determining it in time with a law of nature'. At most, the philosopher's intellectual intuition tells us that some thoughts find their locus entirely within consciousness, but the possibility remains that consciousness itself might yet be causally determined. It remains for Fichte to demonstrate that at the most fundamental level, consciousness itself follows this same form of beginning a state from itself, outside of the determined confines of time. This, of course would be to derive the *freedom* of consciousness, understood on Kantian terms.

In his less philosophically extravagant moments, Fichte is happy to simply appeal to a very crude sense of intellectual intuition which, he maintains, makes clear to anyone capable of it that consciousness *just is*, at the fundamental level, a self-reverting activity, and that those incapable of experiencing this do not deserve a response (*Introductions*, p.20/434). At other, less vitriolic moments Fichte refers to the quality or character of consciousness and appeals to the inability of 'being' to account for such.

Fichte invites us to think of the impossible gaze of the primordial intellect – an intellect that has not ‘discovered’ the concept of I. That is just to say that Fichte invites us to consider the ‘I [that] is supposed to lie there and to have done so before it was grasped in intuition’ (*Sittenlehre* p.37/IV,32). Such an intellect, Fichte thinks, ‘remains a passive onlooker’, an ‘absolute force’ laying dormant (*Sittenlehre* p.37/IV,32). We may say that such an intellect would be blind - there would be no consciousness *of* experience, no experience *for us*, and so, for Fichte, properly speaking, since experience requires the employment of the entire categorial artifice, no experience at all. Instead there would be simply a ‘being’ in the empty sense of the *blind is*.

Effectively, Fichte’s big question then is: by virtue of *what* could a primordial intellect such as this be ‘elevated’ to consciousness such as we are acquainted with? Given that from Kant’s account, objects, space, time and the rest are possible only through the categories, and such a primordial intellect is pre-categorial, by what ability could this passive onlooker become a consciousness such as we would recognize? It cannot be by virtue of ‘things’ or ‘objects’ since these are posterior to this primordial intelligence. Consciousness is possible, Fichte suggests, only by virtue of the ability of the passive onlooker to actively double over and observe itself as observing. This, of course, is precisely that form of activity which was identified in the philosopher’s intellectual intuition, namely self-reverting activity, only now it has been proposed as the original form of consciousness *per se*. As Fichte puts it, it is only thus that ‘[e]yes are inserted into the unified one’ (*Sittenlehre* p.37n/IV,32n). Note then that for Fichte, the act by virtue of which the I intuits itself just is that by virtue of which it is a consciousness.

The I intuits itself purely and simply because it intuits itself
(*Sittenlehre* p.36/IV,32)

If successful, Fichte’s account would accommodate precisely that sense of freedom which Kant located in the thing in itself, but without all the problematic baggage that we saw weighs down Kant’s noumenal appeals. That is, this primordial, pre-reflective intellect, as pre-categorial, is prior to time and space. Time and space, along with

causality and the rest, are nothing *for it*. Rather, only by virtue of its acting on itself does it find itself spatiotemporal. We may say then, that the self-reverting act is *its* first act, the act that gives rise to time and space, and all other consciousness. In this act, and only with this act are time and space ushered in to the picture. Thus time (to butcher Plato) is the moving image of blind indeterminacy becoming determinate consciousness, of the original absolute unity inserting its eyes into itself. In this sense, the act itself is its own beginning of time, and so is not itself preceded by temporal moments. To the extent that all consciousness begins with self-reverting activity then, it seems that the ‘timeless’ criterion set by the Third Antinomy has been met – all conscious agency has its original locus in a timeless, though *non-noumenal* beginning. More, as *self-reverting* act, the consciousness that ensues is the beginning of *its own* state. Thus, the ‘autonomy’ criterion is also met.

§3:i.ii.ii) *Practical Concerns*

Fichte understands that this kind of explanation is not going to convince everyone. It ultimately remains no more than a postulate of theoretical reason. Fichte himself is confident that the philosopher’s intellectual intuition identifies just that *real* initial intuition by which all consciousness is constituted. Indeed, for him it is clear that there simply are no other candidates, and so the case is closed. As he puts it:

Is this original consciousness any different from the one that we, as philosophers have just produced within ourselves? How could it be, given that...the philosopher, as such, certainly possesses no other subjective form of thinking than that common and original form that is present in all reason?

(*Sittenlehre* p.36/IV,31)

Yet, Fichte admits that this is not enough. He has not demonstrated to the ultimate satisfaction of theoretical reason either that self-reverting activity is the fundamental characteristic of consciousness, or that we are transcendently free. Freedom appears to

us a fact of experience, given *eg* the apparently free creation of the concept of an I yielded through the philosopher's intellectual intuition. But Fichte is aware that an appearance of freedom is not, on its own enough for theoretical reason. In true Kantian style though, he adds that it is sufficient for *practical* reason.

the appearance of freedom is an immediate fact of consciousness... [Yet] one might still wish to go further and could thereby transform it into an illusion. There is no theoretical reason for not doing this, but there is a practical one: namely, the firm resolve to grant primacy to practical reason, to hold the moral law to be the true and ultimate determination of our essence, and not to transform it into an illusion by sophistical reasoning
(*Sittenlehre* p.56/IV,53-4)

Recall Allison's attempt to explain away the problems of Kant's moral theory by appealing to practical reason's ability to step into the 'vacant space' where theoretical reason met its tether. It was noted that this move missed the modal distinctness of the intelligible and the thing in itself, and so was subject to all the familiar criticisms of Kant's project. Yet here, note that when Fichte invokes the primacy of practical reason, and more, its ability to fill the empty space of theoretical reason's impasse, there is no modal distinction at play. Rather, freedom and the intelligible are rooted entirely within human experience. Thus, for Fichte, the problems inherent to Kant's invocation of the primacy of the practical that we saw above fail to apply.

Note also that practical reason steps into play just when there is an awareness of self-reverting activity. Whilst, as Fichte admits, sophistical theoretical postulates can be put forward doubting its ultimate origin, *for practical reason*, the awareness of self-reverting activity just is the awareness of our freedom. It is in this sense then, that Fichte comments 'this derivation makes comprehensible the so-called categorical imperative' (*Sittenlehre* p.52/IV,50). Insofar as we take ourselves to be free in our activity, we do so by the law we set ourselves – that we will not go beyond the appearance of freedom, 'which thereby becomes for us the truth' (*Sittenlehre* p.56/IV,54). That is, the very form of

consciousness as a self-reverting activity is that of an 'ought'. It is through this awareness of the fundamentally normative structure of consciousness that the I is an I. To observe oneself as subject in distinction from an object is to notice oneself as bound. It is to think of one's free activity under a law, and a law that one gives oneself. Thus, at the heart of consciousness lies the categorical imperative, a normative law one gives oneself to be autonomous.

If a rational being thinks of itself as self-sufficient...then it necessarily thinks of itself as free, and – this is what really matters here – it thinks this freedom under a law of self-sufficiency. This is the meaning of our deduction.

(*Sittenlehre* p.54/IV,52)

§3:iii) *Intersubjectivity*

To recap then, Fichte attempts to accommodate freedom precisely within the human sphere of meaning from which it had been exiled by Kant. Fichte roots freedom in our conception of ourselves as rational agents, and construes the fundamental form of rationality as normatively loaded. Yet, in explaining morality by appealing exclusively from the conditions of rational agency, might Fichte not fall prey to sorts of concerns about a contorted moral phenomenology, such as were raised against Korsgaard? Fichte thinks that his incorporation of intersubjectivity into his account of individuality will put this concern right.

In the "Second Introduction" Fichte complains that his use of the first-person pronoun (and its negation) has led to grave misunderstandings about the scope of the *Wissenschaftslehre*'s claims heretofore. He specifies that the I of the *Wissenschaftslehre* elucidated above, refers only to the most fundamental structure of subjectivity as taking the form subject/object, and that this elucidates only the 'conditions for all consciousness' (*Introductions* p.61). The subject/object partition of consciousness refers to the mere form of consciousness *per se* (I-hood), and as such, the formal I identified in the concept of an I is *not* to be mistaken for the fully conscious individual.

the I with which the *Wissenschaftslehre* commences...contains nothing but the form of I-hood, [*ie*] self-reverting activity...[which] is not thought of as an individual, because I-hood has not yet been determined as individuality (*Introductions* p.100).

In *Foundations of Natural Right* (hereafter *FNR*) Fichte addresses the place of consciousness of one's individuality within the *Wissenschaftslehre*, and maintains the necessity of intersubjectivity for any such judgement. The deduction in *FNR* begins by rehearsing the deduction of abstract I-hood, here succinctly summarised: Self-reverting act: A, determines its own self as object: B, only by virtue of positing some opposed, limiting object(s): C (*FNR* p.18-9). But it is specified at the very beginning that here the deduction will culminate with the demonstration of the 'finite rational being' (*FNR* p.18). Thus, the deduction of the empirical individual in *FNR* begins with, and builds upon, the deduction of the formal I.

The second stage of the deduction flags up an apparent problem for the preceding tripartite model of the awareness of rational activity. Since this model maintains that awareness of rational activity requires an object to be comprehended as constraining the subject's activity, *ie* as standing in opposition to and limiting the activity of the subject, then it appears that awareness of rational activity has been achieved only by sacrificing the subject's free activity. But of course this cannot be the case, since the object itself 'is supposed to be the subject's own efficacy' (*FNR* p.31). How could it be that the subject is both freely acting, and yet constrained or determined by an object? Thus we are presented 'with a complete contradiction' (*FNR* p.31).

Fichte's proposed solution to the apparent contradiction is to dissolve the problem by reconceptualising the problematic determination as self-determination, born as a response to an invitation to be self-efficacious. Thus, the contradiction is dissolved 'if we think of the subject's being-determined as *its being-determined to be self-determining* i.e., as a summons [*eine Aufforderung*] to the subject, calling upon it to resolve to exercise its efficacy' (*FNR* p.31).

Note that this alone does not necessarily solve, or dissolve the problem of the simultaneity of freedom and determination. If the summons is conceived as an imperative or order, *ie* as the sort of call to which one cannot say no, then the problem remains, as the contradiction between determination and free efficacy still persists.

Fichte then discusses the necessity of the summons as comprehended, or posited as 'given to it from the outside' (*FNR* p.32) confining the investigation only to how the summons 'must appear to the subject under investigation' and *not* to a consideration of it 'from the transcendental view' (*FNR* p.34). The summons must, at the empirical level, be experienced as an external influence on the original 'sheer activity' by calling it to be self-reverting in character.

The next stage of the deduction establishes that due to the purposive character of the summons itself, 'the cause of the summons must itself necessarily possess the concept of reason and freedom' which is just to say that 'it must be an intelligence...and thus a rational being, and must be posited as such' (*FNR* p.35). With this step the aforementioned concern about the summons not necessarily dissolving the contradiction of free efficacy and determination is supposedly relieved. As a summons from a rational individual the subject is free to accept or decline the invitation to act as a free rational being (*FNR* p.33), though of course, even the decision to decline would be to exercise one's free choice, and as such, the realisation of one's individuality is inevitable, given that the summons originates from another rational individual. Thus the deduction is complete. Responding to the summons of another rational agent is necessary in order for me to actualise my own, otherwise latent capacity for full-blown consciousness of myself.

If successful, Fichte has derived the necessity of intersubjectivity as a condition for consciousness of individuality. Thus, our moral acting in the world would be posterior to a genuine interaction with other rational beings. Insofar as we are moral then, we have been summoned to do so by a rational agent. As such, we might think that Fichte, in incorporating intersubjectivity as a condition for moral acting in the world might sidestep

most of the blow accorded to Korsgaard's account. That is, since interaction with other rational beings is required as a pre-cursor for moral actions, perhaps his ethics, though rooted in rational agency might escape the charge of being too subject-oriented in its conception of morality. What is wrong about slavery for Korsgaard is something about my rationality, whereas what is wrong about slavery for Fichte must find its explanation in an actually experienced shared awareness of *us* as rational agents. I suggest though that things might not be so straight forward for Fichte. For Fichte, the conceptual bedrock of the formal I's counterpositing upon which intersubjectivity is based ensures that the encounter with the rational other in conscious experience is reducible to an instance of the counterpositing of subject/object concepts within the subject. In order for there to be any experience of rational beings, on Fichte's model there must be a foundational consciousness, mere I-hood. That is, ultimately, the experience of the other rational individual must, for Fichte, be reducible to a complex conceptual relation, itself a mere fleshing out of the primitive concept of the general 'object' (Not-I). Thus, far from experiencing the other in their immediacy, for Fichte, it seems the other is always experienced at a conceptual remove, as a projection or comprehension of my subjectivity *as* the other. We might say that Fichte remains committed to a *subjective* intersubjectivity, and that thus, what is wrong about slavery, for Fichte as for Korsgaard finds its *ultimate* explanation in a fact about my own rationality.

§3:iv) Fichte and Conscience

It was noted above that conscience seems to play more than a passing role in Kant's own moral thought, and that this presented an obstacle to the constructivists' emphasis on *procedure*. Further, it was noted that Kant himself emphasizes the importance of listening to one's conscience and our 'duty' to strive to ensure it is informed or enlightened within his own moral philosophy. Similarly, Fichte suggests that the command 'Act according to your conscience' is the 'formal condition for the morality of our actions'. So morality is, as much for Fichte as Kant, intricately tied to conscience. Of course the most pressing objection to this line of thought is that surely people are often mistaken in their convictions regarding their duties. Indeed, as we saw, this was the motivation for Kant's

discussion of the duty to inform one's conscience as best we can. It was also noted however, that given Critical ignorance, Kant must be committed to a certain irony within his moral theory – the steps we take to inform conscience may amount to nothing, given the ultimate noumenal grounds of conscience.

To this, Fichte replies that such thought can be taken as an objection to the authority of conscience. Yet, the objection can be admitted only insofar as it could also be admitted that there is 'no absolute criterion for the correctness of my conviction concerning my duty' (*Sittenlehre* p.156/164). If such an absolute benchmark for certainty regarding one's conviction could be identified, then the objection would fall. Note then that where Kant is resigned to a position of Critical ignorance regarding conscience – there is no ultimate assurance that conscience will get it right, due to his unfortunate commitment to its ultimate noumenal source – for Fichte, with the thing in itself out of the picture, the ground is cleared for a *defence* of conscience.

Fichte's argument for the absolute criterion for such correctness begins with the observation that the moral law demands a certain determinate conviction. That is to say, the moral law does not issue in vague terms, but rather commands specific actions or abstentions to be performed for specific reasons, examples of which can be found in Kant's discussions of the potential suicide case at *Gr* 4:422, or the man asked to bear false witness or face execution at *C.Pract.R.* p.30-1/141. As Fichte presciently points out though, a conviction is simply a particular kind of cognition held by rational agents. But the moral law itself is not a cognitive faculty. Thus the moral law is *not* itself a 'power of cognition, and therefore by virtue of its very essence it cannot produce this conviction by itself' (*Sittenlehre* p.157/165). Recall that Kant himself is at pains to distinguish, indeed even distance the theoretical from the practical, though perhaps he might not agree with Fichte that 'the practical power is not a theoretical power' (*Sittenlehre* p.157/165). Nonetheless, Fichte claims that insofar as a conviction can only arise *via* cognition, and the moral law demands a certain conviction, the moral law must be, in some way, related to the theoretical, cognitive power, which Fichte here calls, 'the power of reflective judgement' (*Sittenlehre* p.157/165).

For Fichte, the manner in which the moral law and reflective judgement are related follows from Kant's well-known stance regarding the strictly *formal* character of the moral law. If the moral law is only formal, then whilst it can surely establish the condition of an action or conviction, something else must be invoked to establish *what content* is to be conditioned (see *Sittenlehre* p.145, 155/153, 163). So whilst the moral law determines the form of the conviction, reflective judgement must search for the appropriate matter or content, such that *just this determinate conviction* can be actualized in experience.

It is important to note here that reflective judgement's searching is itself a free act. As Fichte puts the point, '[a]s long as the power of judgement continues to search, the free power of imagination continues to hover between opposites' (*Sittenlehre* p.159/167). To elaborate on this admittedly cryptic remark, Fichte expresses consciousness of the freedom of the self, *qua* formal I, as a consciousness of 'my transition from indeterminacy to determinacy' (*Sittenlehre* p.130/136). The basic thought is, as we have seen, that any determinate object of thought, since it is determinate, is fixed down, and so stands in opposition to the free activity of the self (*Introductions* p.116). Thus, consciousness of my transition from indeterminacy to determinacy is at least a condition for consciousness of myself as, in some sense, that which does the determining whenever there are determinate thoughts (*Sittenlehre* p.131/138). Yet indeterminacy cannot be coherently conceived of as any one specific instance of 'not-determinacy'. That is, for Fichte, it is simply wrong to conceive of indeterminacy as any determinate ' $\neg x$ '. Instead, Fichte conceives of the indeterminacy of the self as the imagination's original, that is, pre-conscious 'undecided hovering between several possible determinations' (*Sittenlehre* p.131/138). Of course, once we actually do think of this or that object of thought, such an object of thought will inevitably be determinate, and thus opposed to the free activity of the self. Yet since any determinate thought could have conceivably been otherwise, freedom, *qua* indeterminate hovering, 'will always be the cause' (*Sittenlehre* p.131/138). Regarding the formal determination of conviction by the moral law then, once reflective judgement gets going with the search for the appropriate material, 'the theoretical powers

persue their own course until they hit upon something that can be approved' (*Sittenlehre* p.157/166 my emphasis). The idea is that whilst the form is determined entirely by the moral law, indeed, that the moral law in some sense sets the imagination off hovering in the first place, the hovering of the imagination is itself a free action, even if such free hovering culminates in the ensuing determination of the material. This is important, since it cements the Kantian link between morality and freedom for Fichte, which in turn entitles him to reject any and all forms of reliance on external authority in moral matters (*Sittenlehre* p.168/176-7). Further though, such a move accommodates freedom for the formal I *even if* the empirical I is determined by the moral law and also helps to explain how the action could be conceived of as *mine* yet subject to determination by the moral law. It is the (formal) I that does the free hovering, even if the end result is a determinate, immediate fact to which the (empirical) I is bound.

Returning to the derivation of the absolute criterion for the correctness of one's conviction regarding duty, we have seen already that even if the moral law determines the form the conviction will take, the power of reflective judgement freely searches for the fitting material so that just this dutiful conviction can be actualized. When, or rather *if*, reflective judgement does find some fitting candidate, then the moral law and theoretical judgement are brought into harmony with each other. Since this results in a determinate duty in experience, the formal I and the empirical I too are brought into harmony with each other, since the free activity of the former is manifested in the empirical activity of the latter; namely, the performance of the moral act. '[T]he original I and the actual I will now be in harmony' (*Sittenlehre* p.158/166-7).

It is important to note here the parallel between Fichte's choice of the terms 'reflective judgement' and Kant's thoughts on reflective judgement in *C.ofJ*. I take it that the idea is that the harmony in the experience of the moral law and the theoretical power of imagination's free hovering is importantly similar to an aesthetic experience – once harmony is achieved, there is a 'repose', and a state of 'peace' (*Sittenlehre* p.140/147). Note though that this peaceful repose is to be distinguished from an aesthetic experience in that no pleasure is gained by it. Instead, this 'cold approval' is merely experienced as a

‘feeling of harmony’ (p.159/167). Yet as with aesthetic experience, which we experience *immediately*, though only by virtue of the interplay of sensibility and understanding brought to the fore by transcendental inquiry, similarly here, what has been derived at the transcendental level as the reciprocal interaction of the moral law or practical power with reflective judgement or theoretical power, is felt in experience as an ‘*immediate certainty*’ (*Sittenlehre* p.159/168). As soon as the relevant material is found by reflective judgement, then ‘constraint is present’ and we, as empirical subjects are determined by the law to form some determinate conviction.

As soon as the power of judgement finds what was demanded, the fact that this is indeed what was commanded reveals itself through a feeling of harmony...I cannot view this matter in any way other than in the way I do view it: constraint is present, as it is in the case of every feeling. This feeling provides cognition with *immediate certainty*, with which *calm* and satisfaction are connected.
(*Sittenlehre* p.159/168-9)

it is not by means of argumentation that I know whether I am in doubt or certain...instead, this is something I know only through immediate feeling. Only in this way can we explain subjective certainty as a state of mind. The feeling of certainty, however, is always an immediate harmony of our consciousness with our original I – nor could things be otherwise in a philosophy that begins with the I.
(*Sittenlehre* p.161/169).

This immediate certainty about our determinate duty, which is, for Fichte, necessarily non-discursive, but rather *felt*, is just what he means by ‘conscience’. ‘Conscience is ... *the immediate consciousness of our duty*’ (*Sittenlehre* p.164/173). It is precisely this *feeling* of immediate certainty that he thinks provides the absolute criterion for the correctness of conviction regarding duty that we started with. We know it is our duty because we experience the harmony of freely reflective judgement and the moral law in our empirical experience of the world, and we experience it as an immediate non-

conceptual feeling that we need look no further for the right thing to do – it is simply immediately clear to us that we ought to act in just this way at just this moment. Note though, that although our determinate duty is immediately certain to us in experience, it is premised on the free acting of the imagination. That is, once given, or found, the consciousness that *just this material 'x'* is the content of my duty, is immediately certain. Yet the material itself is found only through *my* act of *freely* reflective thought, a hovering between indeterminacies. Thus, for Fichte, whilst the morally right action is immediately and unavoidably felt by one's conscience, it does not *originate* in one's conscience, but rather in one's own freedom. Whilst the empirical I cannot but feel the determination of its conviction by the moral law in conscience, that the moral law is efficacious on his conscience is only possible as a consequence of the exercise of the formal I's free activity. So while the empirical I feels the command of the moral law *immediately*, the attribution of *freedom* to the formal I enables a reconciliation of the two otherwise contrary notions.

Conclusion.

In his *Problems of Moral Philosophy* Adorno comments:

if we wished to summarize the distinction between Kant and Fichte and all succeeding idealists in a single gesture it would be this moment of rupture, the moment when Kant says, 'This need not concern us', that they were unable stomach. And what they said was, 'What you have said about this being of no concern to us is precisely what concerns us most.'

(Adorno (2001) p.95)

As I hope to have shown, this remark seems, in any ways, an apt description of Fichte's moral project. Where Kant tells us that freedom must remain unknowable, Fichte derives an account of freedom. Where Kant tells us we cannot know how conscience operates, Fichte offers a derivation of conscience. Yet, Fichte's moral philosophy is unlikely to find adherents. At best we may say that his account of conscience is coherent, given his

other commitments. At worst, we might view it as among the worst forms of speculative excess. More, his account of the intersubjective nature of individuality, and thereby also the intersubjective element underlying all our moral consciousness, is, I suggest a failure. Fichte's account is plagued here by the all-pervading account of foundational consciousness, so that the other always remains at best a 'posited' rational being. Yet, I suggest that Fichte's account of freedom is, by Kantian strictures, a success. He manages to reconcile freedom with the sphere of human experience and meaning in a way that Kant never could, indeed, in conscious awareness of the limitations of the Kantian system's use of the thing in itself. More, though his account of conscience is, I suggest, unconvincing, Fichte's insistence that 'pay[ing] attention to the dictates of one's conscience... is sufficient for engendering both a dutiful disposition and dutiful conduct' (*Sittenlehre* p.20/IV,14) places him firmly on the same page as Kant with regard the 'state of play' of morality both in experience, and within philosophy. Both Kant and Fichte share the thought that ordinary, pre-philosophical reasoning needs little by way moral education. More, they share the thought that philosophy ought to remain true to this awareness. Fichte's failures then, are explicable by their proximity to the central themes of Kant's project. His success though is a tribute to his keen awareness of the inappropriateness of the thing in itself for Kant's moral philosophy.

Bibliography.

Adorno, T.W. (2001) *Problems of Moral Philosophy* (Stanford University Press: Stanford, California)

Allison, H. (1986) "Morality and Freedom: Kant's Reciprocity Thesis" *Philosophical Review* 95

- (1990) *Kant's Theory of Freedom*. (CUP: Cambridge)

Ameriks, K. (2000) *Kant and the Fate of Autonomy: Problems in the Appropriation of the Critical Philosophy*. (CUP: Cambridge)

Bird, G. (2006) *The Revolutionary Kant: A Commentary on the Critique of Pure Reason*. (Open Court: Chicago)

Coplestone, F. (1965) *From the Post Kantian Idealists to Marx, Kierkegaard, and Nietzsche*. Vol. 7 of *A History of Philosophy* (Doubleday: New York).

DiGiovanni, G. (2005) *Freedom and Religion in Kant and his Successors: The Vocation of Humankind, 1774-1800* (CUP: Cambridge)

Fichte, J.G. (1988) *Fichte: Early Philosophical Writings* Breazeale, D. (trans. And ed.) (Cornell: Ithaca)

- (1994) *Introductions to the Wissenschaftslehre*. (Hackett: Indianapolis/Cambridge)

- (2000) *Foundations of Natural Right* Neuhauser, F. (ed), Baur, M. (trans) (CUP: Cambridge)

- (2005) *The System of Ethics According to the Principles of the Wissenschaftslehre*, Breazeale, D. & Zöller, G. (trans. & ed.) Cambridge University Press: Cambridge.

- Förster, E. (1989) (ed.) *Kant's Transcendental Deductions: The Three 'Critiques' and the 'Opus postuum'* (Stanford University Press: Stanford)

- (2012) *The Twenty-Five Years of Philosophy: A Systematic Reconstruction*. (Harvard University Press: Cambridge, Mass.)

- Freeman, S. (2003) *The Cambridge Companion to Rawls* (CUP: Cambridge)

- Freyenhagen, F.
 (2008) "Reasoning Takes Time: On Allison and the Timelessness of the Intelligible Self" in *Kantian Review* 13 (2): 67-84

- Gardner, S. (1999) *Kant and the Critique of Pure Reason*. London: Routledge.

- Geiger, I. (2010) "What is the Use of the Universal Law Formulation of the Categorical Imperative?" in *British Journal for the History of Philosophy*, 18:2, 271-295.

- Kant, I. (1998) *Critique of Pure Reason*. Guyer, P. & Wood, A. (ed. & trans.) (CUP: New York)

- (1950) *Prolegomena to any Future Metaphysics* Beck, L.W. (trans.) (Bobbs-Merrill: Indianapolis)

- (1998) *Groundwork of the Metaphysics of Morals* Gregor, M. (ed & trans) (CUP: Cambridge)
- (2004) *Critique of Practical Reason*. Abbott, T. K. (trans) (Dover: New York)
- (2007) *Critique of Judgement*. Meredith, J. C. (trans) (OUP: Oxford).
- (1991) *The Metaphysics of Morals*. Gregor, M. (trans) (CUP: Cambridge)

Korsgaard, C. (1996) *Creating the Kingdom of Ends* (CUP: Cambridge).

- (1996) *Sources of Normativity* (CUP: Cambridge)

Kosch, M. (2006) *Freedom and Reason in Kant, Schelling and Kierkegaard*. (OUP: Oxford).

Martin, W. (1997) *Idealism and Objectivity: Understanding Fichte's Jena Project* (Stanford University Press: Stanford, California)

O'Neill, O. (1989) *Constructions of Reason: Explorations of Kant's practical philosophy*. (CUP: Cambridge).

Paton, H.J. (1946) *The Categorical Imperative* (Hutchinson's University Library)

Rawls, J. (1975) "The Independence of Moral Theory" in *John Rawls: Collected Papers* (1999) Freedman, S. (ed) (Harvard University Press: Cambridge, Mass.)

- (1980) 'Kantian Constructivism in Moral Theory' in *The Journal of Philosophy*, September, 1980

- (1993) *Political Liberalism* (Columbia University Press: New York)
- (1999) *A Theory of Justice*. (Belknap, Harvard University Press: Cambridge, Mass.)

Schopenhauer, A.

(1958) *The World as Will and Representaion: Volume 1* Payne, E.F.J
(trans) (Dover: New York)

Watkins, E and Fitzpatrick, W.

(2002) ‘O’Neill and Korsgaard on the Construction of Normativity’ in
The Journal of Value Inquiry **36**: 349-367.

Wood, A. (1984) “Kant’s Compatibilism” in Wood, A. (ed) *Self and Nature in Kant’s Philosophy* (Cornell University Press: Ithaca)

- (2008) “Kant’s Formulations of the Moral Law” in Bird, G. (ed) *A Companion to Kant* (Wiley-Blackwell: Chichester).